

**UNIVERSIDAD COMPLUTENSE DE MADRID**  
**FACULTAD DE CIENCIAS DE LA INFORMACIÓN**



**TESIS DOCTORAL**

**Periodismo robot.**  
**Aplicaciones de la tecnología, los algoritmos y la**  
**automatización en la comunicación**

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

**Adrián Blanco Ramos**

Directora

**Pedro García-Alonso Montoya**  
**Rafael Carrasco Polaino**  
**Alfonso Javier Fernández del Moral**

Madrid

**UNIVERSIDAD COMPLUTENSE DE MADRID**  
**FACULTAD DE CIENCIAS DE LA INFORMACIÓN**



**TESIS DOCTORAL**

Periodismo robot.

Aplicaciones de la tecnología, los algoritmos y la automatización en la comunicación.

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Adrián Blanco Ramos

DIRECTORES

Pedro García-Alonso Montoya  
Rafael Carrasco Polaino  
Alfonso Javier Fernández del Moral

Washington D.C., 2022



**UNIVERSIDAD COMPLUTENSE DE MADRID**  
**FACULTAD DE CIENCIAS DE LA INFORMACIÓN**



**TESIS DOCTORAL**

Periodismo robot.

Aplicaciones de la tecnología, los algoritmos y la automatización en la comunicación.

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Adrián Blanco Ramos

DIRECTORES

Pedro García-Alonso Montoya  
Rafael Carrasco Polaino  
Alfonso Javier Fernández del Moral

Washington D.C., 2022





## **Agradecimientos**

A mis padres, Cheli y Tomás, y a Mariel, sin cuyo apoyo no podría haber escrito ni una palabra de esta tesis.



## ÍNDICE DE CONTENIDOS

<b>ABSTRACT</b>	11
<b>KEYWORDS</b>	13
<b>RESUMEN</b>	15
<b>PALABRAS CLAVE</b>	17
<b>INTRODUCCIÓN</b>	21
I. Objeto de la investigación	23
II. Estado de la cuestión	27
III. Metodología de trabajo	35
IV. Fuentes de investigación	39
V. Hipótesis de partida	43
<b>CAPÍTULO I: Los inicios</b>	47
1.1. Un primer intento para predecir el resultado de una elección	49
1.2. Periodismo de precisión	55
1.3. Periodismo asistido por ordenador (CAR)	60
1.4. La expansión del periodismo de datos	62
<b>CAPÍTULO II: Periodismo robot</b>	65
2.1. Periodismo robot como término global	70
2.2. Retrato robot de los nuevos profesionales de la información	71

<b>CAPÍTULO III: Automatización en la redacción</b>	<b>77</b>
3.1. Beneficios de la automatización	80
3.2. Riesgos de la automatización	81
3.3. La cobertura de elecciones	82
3.4. Automatización en forma de robots	86
3.4.1. Bots de uso interno	88
3.4.2. Bots públicos	92
3.5. Inteligencia artificial y Newsmaking	100
<b>CAPÍTULO IV: Automatización en el periodismo de investigación</b>	<b>105</b>
4.1. Análisis e investigación de algoritmos	109
4.1.1. Un caso de éxito: <i>The Markup</i> y la investigación de algoritmos	114
4.2. Algoritmos al servicio del análisis de datos	124
4.2.1. Análisis de datos en la investigación <i>The Opioid Files</i>	125
4.3. El uso de sensores en el periodismo de investigación	130
4.4. Machine learning aplicado al periodismo de investigación	140
<b>CAPÍTULO V: Automatización en periodismo visual</b>	<b>143</b>
5.1. El papel de los lenguajes de programación en el periodismo visual	147
5.2. Herramientas automatizadas para la creación de visualizaciones	152
5.3. Automatización y periodismo visual en tiempos de pandemia	157
5.4. Humanos y máquinas para cubrir la pandemia, el caso de <i>The COVID Tracking Project</i>	171
5.5. Automatización y bases de datos propias, la estrategia de <i>The New York Times</i> en la cobertura de la pandemia de la COVID-19	177
5.6. Simulaciones y juegos, el papel de la automatización	182

<b>VI. CONCLUSIONES</b>	192
<b>VII. REFERENCIAS BIBLIOGRÁFICAS</b>	205
<b>VIII. APÉNDICES</b>	230
Relación de siglas, acrónimos y abreviaturas	232
Glosario	235
Índice de tablas	238
Índice de figuras	239
<b>IX. ANEXOS</b>	243
Anexo I. Formulario para entrevistas a profesionales de la información sobre su trabajo de análisis y la visualización de datos de la COVID-19	245
Anexo II. Artículo preliminar del estudio <i>Effectiveness, clarity and impact of visual journalism in the time of a pandemic</i>	254
Anexo III. Trackers seleccionados para el estudio de la automatización en la cobertura de la COVID-19	262
Anexo IV. Selección de bots con fines periodísticos analizados en esta tesis	288



## **Abstract**

This dissertation exposes and analyzes how technology, the use of algorithms and automation — techniques from disciplines such as Math, Statistics or Computer engineering — have transformed journalism and the role of the reporter.

This study creates a taxonomy of the different disciplines that have emerged from the intersection between journalism and technology, describing their main characteristics, evolution over time, benefits and challenges journalists face today working in this field.

In addition, this thesis investigates the impact that automation has had in journalism, analyzing the future of mass media under the umbrella of robot journalism.



## **Keywords**

Robot journalism, automation, algorithms, computation, data



## **Resumen**

Esta disertación expone y analiza cómo la tecnología, el uso de algoritmos y la automatización de procesos — técnicas propias de disciplinas como la matemática, la estadística o la ingeniería informática — han transformado el sector del periodismo y el perfil del profesional de la información.

Este estudio elabora una taxonomía y describe las diferentes disciplinas periodísticas que han surgido en la intersección entre periodismo y tecnología, así como sus características, evolución, beneficios y desafíos que suponen.

Además, esta tesis investiga el impacto que la automatización, es decir, el periodismo robot, ha tenido en el periodismo, analizando el futuro de la comunicación bajo el paraguas del periodismo robot.



## **Palabras clave**

Algoritmos, automatización, periodismo robot, computación, datos



# **INTRODUCCIÓN**



## I. Objetivos de la investigación

**A**utomated journalism o periodismo automatizado por su traducción al español, *algorithmic journalism* o periodismo de algoritmos, *sensor journalism* o periodismo de sensores, *data journalism* o periodismo de datos, *robot journalism* o periodismo robot, *computational journalism* o periodismo computacional son algunas de las especialidades acuñadas en las dos últimas décadas en el sector del periodismo.

El afán por conceptualizar este amplio rango de técnicas y especialidades, surgidas en la intersección del periodismo y la tecnología, persigue reaccionar a las nuevas realidades del periodismo moderno y definir las nuevas tareas que los periodistas, infografistas, diseñadores o ingenieros que trabajan en un medio de comunicación realizan en su día a día.

La incorporación de la tecnología en el sector del periodismo ha abierto nuevos caminos para hacer periodismo y ha dotado de nuevas técnicas al periodista para investigar contenidos e informar al lector.

Los conceptos anteriores son, entre otros, algunos de los términos que los académicos e investigadores utilizan en la literatura para definir diferentes procesos tecnológicos aplicados al oficio del periodismo. En numerosas ocasiones, la academia recurre a estos conceptos para describir realidades similares ya que estas especializaciones comparten las mismas bases.

Entre los fundamentos en común de estas conceptualizaciones se encuentran el uso de los datos como la base de las informaciones o de las tareas periodísticas, así como la aplicación de los lenguajes de programación y la tecnología para manipular, analizar o visualizar, entre otras acciones, esos mismos datos. Las responsabilidades de los periodistas van desde la investigación y redacción de informaciones hasta la colaboración, desarrollo, incluso automatización de otras tareas en la redacción de un medio de comunicación.

Aunque la frontera entre unas especializaciones y otras es difusa, y una buena parte de los términos mencionados anteriormente se utiliza de forma intercambiable, cada uno de estas técnicas engloba matices diferentes

Esta tesis se centra en el estudio de la relación entre el periodismo, la tecnología y la automatización. Por ello, este estudio plantea el estudio del periodismo robot, un término que se usa en esta disertación con el objetivo de englobar las especializaciones citadas. El motivo principal por el que se utiliza el concepto “periodismo robot” es facilitar la comprensión de la aplicación de los nuevos procesos tecnológicos de computación en el sector del periodismo.

De esta forma, en este estudio, se entiende por periodismo robot cualquier aplicación de la tecnología en el día a día de un profesional de la información, desde el uso de lenguajes de programación para facilitar la creación de historias visuales hasta la automatización de procesos para realizar investigaciones o mejorar tareas en una redacción.

Para verter luz sobre el periodismo robot y el resto de las especialidades periodísticas que giran en torno a los datos, la tecnología y la automatización, esta tesis analiza casos de

éxito e iniciativas innovadoras llevadas a cabo por diferentes medios de comunicación a nivel mundial.

Dentro de este enfoque global que gira en torno a la intersección del periodismo y la tecnología, esta disertación plantea una serie de capítulos y secciones como son la automatización de procesos para mejorar las tareas diarias en una redacción, el uso de datos y algoritmos para contar historias que de otra forma no podrían ser contadas y la evolución del periodismo visual gracias a las nuevas tecnologías.

Las principales preguntas que esta investigación pretende responder y que han guiado su confección y estructura narrativa, siguen a continuación:

- ¿Qué es periodismo robot? ¿Cuáles son sus características?
- ¿Qué efectos ha tenido la tecnología en el periodismo?
- ¿Qué posibilidades, oportunidades y desafíos ofrece la tecnología al sector del periodismo y a las labores diarias del periodista?
- ¿Qué impacto ha supuesto la automatización de procesos y el uso de algoritmos en el sector del periodismo?
- ¿Qué efecto ha tenido la pandemia global de la COVID-19 en el uso de las tecnologías en sector del periodismo?
- ¿Cuál es el futuro del periodismo robot y la presencia de la tecnología en los medios de comunicación?

Sección a sección se responderá a estas y a otras muchas preguntas, aclarando algunos de los conceptos más técnicos con ejemplos de reportajes, iniciativas y procesos desarrollados en el sector del periodismo durante el último medio siglo.



## II. Estado de la cuestión

**D**e una forma u otra y con un menor o mayor impacto, la tecnología y la automatización de procesos, en definitiva, el periodismo robot, ha cambiado las costumbres y quehaceres tanto en medios de comunicación pequeños y locales como en medios grandes y globales.

En la última década, la mayoría de los grandes medios de comunicación internacionales ha iniciado o creado algún proyecto de automatización estratégico. Estos proyectos cuentan con un componente de inteligencia artificial con el fin de contar historias o mejorar las labores diarias en una redacción. El sector del periodismo local también ha desarrollado este tipo de procesos a menor escala, a pesar de las dificultades y la falta de recursos que muchos de estos medios afrontan.

*Bloomberg News*, el gigante de noticias estadounidense cuya cobertura mediática se centra principalmente en la economía y los mercados de valores, utiliza la tecnología para

automatizar alrededor de un tercio del contenido generado en su web <sup>1</sup>. Uno de los sistemas más populares que utiliza Bloomberg para la automatización de contenidos ha sido bautizado como Project Cyborg <sup>2</sup>. Este programa de inteligencia artificial asiste a los reporteros a convertir los resultados trimestrales de las empresas en historias, esbozando titulares y seleccionando los contenidos más relevantes que un periodista incluiría en una pieza de acuerdo con la estructura de la pirámide invertida <sup>3</sup>.

A día de hoy, las iniciativas de automatización de procesos y aplicación de la tecnología en la creación de contenidos son numerosas en el sector del periodismo. Una de las aplicaciones de la tecnología en el periodismo más manida por los medios de comunicación es la creación de artículos de manera automática. En la mayoría de los casos, esta técnica ha sido empleada para generar crónicas deportivas, económicas y electorales.

El periódico estadounidense *The Washington Post* experimentó en 2016 con la generación automática de crónicas de partidos de competiciones deportivas escolares <sup>4</sup>. Los profesionales de la información del diario estadounidense dieron vida a un robot de creación de contenidos que denominaron *Heliograf*.

---

<sup>1</sup> Vildrim, P. (9 de abril de 2019). *Will ai save journalism -- or kill it?* Knowledge at Wharton. Consultado a 5 de mayo de 2020, disponible en: <https://knowledge.wharton.upenn.edu/article/ai-in-journalism/>

<sup>2</sup> Ibid

<sup>3</sup> Mullin, B. (21 de mayo de 2016). *Bloomberg EIC: Automation is 'crucial to the future of journalism'*. Poynter. Consultado a 2 de febrero de 2019, disponible en: <https://www.poynter.org/tech-tools/2016/bloomberg-eic-automation-is-crucial-to-the-future-of-journalism/>

<sup>4</sup> The Washington Post. (17 de noviembre de 2020) *High School Football this week: James Madison at South Lakes*. The Washington Post. Consultado a 5 de abril de 2021, disponible en: <https://www.washingtonpost.com/allmetsports/2017-fall/games/football/91619/>

Unos meses más tarde, los periodistas de *The Washington Post* aplicarían las lecciones aprendidas con el robot *Heliograf* para producir crónicas breves con los resultados de elecciones a nivel estatal y regional <sup>5</sup>.

En 2014, la agencia de noticias *Associated Press (AP)* utilizó también técnicas de inteligencia artificial para expandir la cobertura resultados de hasta 4.000 compañías <sup>6</sup>. Reuters, por su parte, ha desarrollado proyectos basados en la inteligencia artificial para generar cientos de historias en múltiples idiomas <sup>7</sup>.

Pero el uso de la tecnología, la automatización de procesos y el desarrollo de aplicaciones de inteligencia artificial no es un fenómeno único del continente norteamericano o de los grandes medios. A nivel nacional, el diario digital español *El Confidencial* también ha experimentado con la redacción automática de crónicas sobre partidos de fútbol <sup>8</sup>.

Este proceso de automatización es efectivo para la generación de textos y piezas que mantienen una estructura común, repetitiva y reproducible. Es decir, una crónica deportiva normalmente incluye el resultado, los anotadores o los jugadores más destacados del partido. Por su parte, una crónica diaria de la bolsa suele destacar las compañías cuyo valor ha crecido más en el mercado de valores, el precio de la acción y los valores cuyo precio de la acción se ha reducido más.

---

<sup>5</sup> The Washington Post. (6 de noviembre de 2018). *Sen. Dianne Feinstein wins California Senate Seat*. The Washington Post. Consultado a 5 de abril de 2021, disponible en: <https://www.washingtonpost.com/election-results/california-senate-class-1/>

<sup>6</sup> Blankespoor, E., Dehaan, E. y Zhu, C. (10 de mayo de 2017). *Robo-journalism and capital markets*. Stanford Graduate School of Business. Consultado a 5 de abril de 2021, disponible en: <https://www.gsb.stanford.edu/faculty-research/working-papers/robo-journalism-capital-markets>

<sup>7</sup> Stray, J. (2016). *The age of the Cyborg*. Columbia Journalism Review. Consultado a 7 de abril de 2021, disponible en: [https://www.cjr.org/analysis/cyborg\\_virtual\\_reality\\_reuters\\_tracer.php](https://www.cjr.org/analysis/cyborg_virtual_reality_reuters_tracer.php)

<sup>8</sup> Rojas, J. L. y Toural, C. (2019). *Automated sports journalism. The AnaFut case study, the bot developed by El Confidencial for writing football match reports*. Doxa Comunicación, (29), 235-254. Disponible en: [https://repositorioinstitucional.ceu.es/bitstream/10637/10745/2/en\\_m4\\_stamped.pdf](https://repositorioinstitucional.ceu.es/bitstream/10637/10745/2/en_m4_stamped.pdf)

En ambos casos, esa previsibilidad permite que el texto final pueda escribirse con antelación, incluyendo algunas alteraciones y condicionales, con el fin de automatizar el proceso de redacción por parte de un robot. Otra opción, más compleja técnicamente y que supone el uso de técnicas de inteligencia artificial, es dejar al robot que aprenda las bases de un artículo sencillo y lo desarrolle por sí mismo.

Al cierre de la bolsa o al finalizar un partido, el robot puede actualizar los valores y nombres propios de la crónica y completar la noticia con la información nueva. El número de condicionales que se establezcan en el programa y la información de entrenamiento para el aprendizaje del robot, determinarán la calidad del texto, la cohesión y la variedad en el uso del lenguaje.

Otra iniciativa, en este caso enfocada a interactuar con el lector, es la apuesta del medio digital español *Maldita*, un medio sin ánimo de lucro que combate la desinformación mediante técnicas de verificación de hechos (fact-checking), por un chatbot en la plataforma social WhatsApp.

En el sector del periodismo, un chatbot es un robot conversacional que tiene la capacidad de mantener una conversación digital con el lector y ofrecer información. Mediante el desarrollo de uno de estos robots, *Maldita* automatizó un canal de WhatsApp para informar al lector sobre noticias falsas <sup>9</sup>.

La automatización en la creación de artículos de información periodística o la interacción con el lector de forma automática son sólo algunas de las aplicaciones de la tecnología y de la automatización en el campo del periodismo.

Aunque la academia y la literatura tienden a usar el término periodismo robot para este tipo de casos, lo cierto es que los robots — los archivos o documentos de código escritos

---

<sup>9</sup> Maldita.es. (Junio de 2021). *Disinformation on Whatsapp: maldita.es chatbot and the frequently forwarded attribute*, Maldita. Consultado a 15 de Enero de 2022, disponible en: [https://maldita.es/uploads/public/docs/disinformation\\_on\\_whatsapp\\_ff.pdf](https://maldita.es/uploads/public/docs/disinformation_on_whatsapp_ff.pdf)

por un periodista o profesional de la información con el fin de ejecutar una tarea — forman parte de realidades más complejas, así como de disciplinas más amplias como la investigación, el análisis de datos o la visualización de datos, entre otras. De hecho, los reportajes de investigación sobre cómo los algoritmos creados por terceros — gobiernos, administraciones o compañías privadas — influyen en la vida de las personas, comienzan a hacerse hueco entre las noticias de portada de los medios a nivel internacional <sup>10</sup>.

Por tanto, esta tesis propone el uso del término periodismo robot para abarcar un campo más amplio, no sólo limitado a la creación de piezas periodísticas de forma automática.

¿A qué se debe esta propuesta? En la mayoría de las ocasiones, un proceso no se automatiza de principio a fin. Pese al valor que aportan los robots mencionados, como Cyborg o Heliograf, estos no son perfectos y pueden cometer errores o producir textos planos sin mucho recorrido. En la mayoría de los casos, es más efectivo automatizar pequeños procesos que ayuden al periodista a realizar su trabajo en otros campos periodísticos, como la visualización de datos o la investigación. Dichos procesos no suelen ser reconocidos al utilizar el término periodismo robot o otro concepto, como periodismo de datos.

A diario, los periodistas de datos, periodistas visuales, diseñadores e ingenieros contribuyen al campo de la automatización, o en este caso del periodismo robot, cuando escriben un programa o desarrollan un robot que analiza una serie de datos, genera un gráfico o crea una nueva experiencia para presentar la información. En el momento de escribir código informático, el periodista establece y ordena una serie de instrucciones a un ordenador (robot) para que éste las interprete. El ordenador procesa estas instrucciones y las traduce al lenguaje binario, interpretándolas y mostrando el resultado esperado.

Por ello, esta tesis abarca y analiza la influencia que ha tenido la tecnología, la computación, la automatización, los algoritmos y, en definitiva, los datos, en el periodismo.

---

<sup>10</sup> El diario español El País, entre otros medios de comunicación, cuenta con una sección titulada Algoritmos Computacionales, disponible en <https://elpais.com/noticias/algoritmos-computacionales/>

De esta forma, se ha elegido el término periodismo robot para referirse a cualquier proceso informático que automatiza, apoya, ayuda y mejora el proceso periodístico.

Con el fin de definir los diferentes conceptos que forman parte de la vertiente más tecnológica del periodismo actual, este trabajo analiza las diferentes especializaciones periodísticas que se enmarcan en la intersección entre el periodismo y la tecnología, mostrando sus similitudes y sus diferencias.

Ante la variedad de elementos, conceptos y disciplinas que han surgido en la intersección del periodismo y la tecnología en los últimos años, esta tesis pretende analizar la salud de la que gozan estas disciplinas y los desafíos que afrontan sus profesionales en el mundo digital.

La multitud de especializaciones que han surgido en los últimos años demuestra las oportunidades que ofrece la tecnología para mejorar las coberturas periodísticas. En cualquier caso, la pervivencia en el tiempo de estas disciplinas noveles dependerá de que se sigan practicando en el tiempo, sus contenidos sean relevantes para el lector y su ejecución sea sostenible en relación con el uso de recursos y con el número de profesionales con conocimientos técnicos para desarrollarlas.

La tecnología ha cambiado los procesos periodísticos y ha hecho que los profesionales de la información adquieran otros conocimientos y se especialicen en disciplinas complementarias al oficio del periodismo. Esta tesis analiza también cómo estos periodistas, representados en su mayoría por el periodista de datos, se han ido incorporando e integrando en las redacciones de los medios de comunicación de diferentes partes del planeta.

Además, esta disertación examina cómo han evolucionado dos pilares fundamentales del periodismo como son el periodismo de investigación y el periodismo visual gracias a la incorporación de estas nuevas tecnologías. Ente otras cosas, explora qué puertas se han abierto para el oficio periodístico en la nueva realidad tecnológica. También expone cuáles son los desafíos que se le plantean al periodismo — un sector establecido desde el siglo

XVIII con la publicación de los primeros diarios de prensa y, por tanto, normalmente conservador y reticente al cambio — ante una adopción tecnológica tan rápida en apenas un par de décadas.

La evolución veloz del periodismo robot, en apenas una década, y la constante aplicación de nuevas tecnologías en el periodismo, puede verse reflejada en la literatura académica, donde se han definido y estudiado nuevas disciplinas que tienen en común la combinación del periodismo y de la tecnología. Ante los cambios constantes y trepidantes que se han producido en el sector del periodismo en los últimos años, esta tesis pretende estudiar los más significativos y fijar un marco para el estudio de los procesos de automatización, el uso de algoritmos, los lenguajes de programación y la inteligencia artificial. En definitiva, esta tesis pretende diseccionar la tecnología que los medios de comunicación y sus periodistas utilizan y ponen en práctica en el día a día.



### **III. Metodología de trabajo**

**P**ara la elaboración de esta tesis doctoral se ha seguido una metodología analítica y descriptiva de las diferentes disciplinas y tecnologías que el periodismo ha utilizado y utiliza tanto en el pasado como en el presente. Para ello, se han tenido en cuenta las reflexiones y exposiciones de otros autores y las referencias en la literatura sobre el periodismo a lo largo de los últimos sesenta años.

Con el fin de mantener una coherencia y una cohesión en esta tesis, el término periodismo robot ha sido el concepto elegido para referirse de forma global y general a todas y cada una de estas disciplinas que comparten y aplican el uso de la tecnología y los datos en el periodismo.

El repaso histórico a este tipo de disciplinas, tecnologías y técnicas es importante para aprender cómo los periodistas han interiorizado diferentes conocimientos y técnicas que les han permitido dar un paso más allá a la hora de cubrir las historias del día a día. Con el paso de los años, la tecnología, los datos y la automatización están más presentes en las redacciones. Por ello, para entender el denominado periodismo robot en esta tesis, hay que comprender primero el papel y la evolución de la tecnología en el sector del periodismo.

Debido a la presencia de tecnicismos y realidades ajenas al periodismo clásico, esta tesis pretende ser muy descriptiva, definiendo términos y explicando en detalle las técnicas, tareas y procesos que los profesionales de la información realizan en su día a día.

Es por ello que, además de las amplias explicaciones que se incluyen en esta tesis sobre los tecnicismos utilizados a lo largo de sus páginas, también se ha incluido un glosario al final de la misma para ampliar los conceptos mencionados y expuestos en el texto.

Por el carácter técnico de esta tesis y el enfoque en las tareas tecnológicas que se llevan a cabo en una redacción, la lectura de documentación técnica sobre la sintaxis de los lenguajes de programación, las librerías de código o los programas de software ha sido fundamental para comprender el rol y las responsabilidades del periodista actual. La documentación técnica, que detalla los diferentes procesos tecnológicos analizados en esta tesis, ha sido clave para entender cómo trabajan los periodistas de datos, sus tareas y responsabilidades en la intersección del periodismo y la tecnología.

Por otro lado, la comprensión de la documentación de los softwares o lenguajes de programación más comunes ha sentado las bases para explicar y definir los diferentes conceptos y procesos más complejos y técnicos presentes en esta tesis. Por esta razón, esta disertación incluye numerosos enlaces a GitHub, la plataforma para almacenar código por excelencia, así como a las páginas web oficiales de diferentes lenguajes de programación y librerías de código.

Con el fin de analizar qué técnicas han abierto nuevas puertas a la labor periodística, a la investigación y a la presentación de esas historias, se han analizado decenas de artículos relevantes publicados durante los últimos años en los que se ha puesto en práctica, entre otras, alguna de las siguientes prácticas:

- Investigación de algoritmos.
- Creación de algoritmos.

- Automatización de procesos.
- Uso de sensores para la recolección.
- Uso de lenguajes de programación para analizar o visualizar datos.
- Aplicación de técnicas de inteligencia artificial.

Entrevistas *on the record* y *off the record*, presentaciones, charlas, talleres han servido de base para conocer y describir los detalles de los diferentes proyectos de automatización descritos en esta tesis. Sin dichas conversaciones formales e informales hubiese sido imposible conocer algunos de los detalles mejor guardados obre la ejecución de estaos proyectos debido a que el código y diseño de estas ideas no siempre es público y está disponible para consulta.

Con el fin de trabajar de la forma más estructurada posible, las notas de dichas interacciones con otros profesionales de la información han sido organizadas por contenidos para poder combinar experiencias con lecturas de libros, artículos académicos o análisis en publicaciones especializadas sobre el periodismo y la tecnología.

Con el objetivo de tratar la información de la forma más objetiva posible, la información se ha estructurado en listas y bases de datos cuando ha sido posible. Para esta tesis se han creado una serie de bases de datos de artículos, piezas periodísticas y tareas donde se ha aplicado la tecnología o la automatización en una redacción periodística. La creación de una base de datos de contenidos específica para el estudio de estos fenómenos ha sido de gran utilidad para seleccionar la información más relevante, identificar los contenidos más innovadores y tratar la información de la forma más justa y neutral posible.

Por último, mi experiencia como periodista en el periódico estadounidense *The Washington Post* también ha jugado un papel fundamental para conocer cómo otros reporteros aplican la tecnología para realizar su trabajo. El trabajo desarrollado en lo últimos años en el diario estadounidense me ha permitido conocer de primera mano cómo se usa la tecnología en una redacción de noticias, además de desarrollar y aplicar el periodismo que defino en esta tesis en diferentes tareas y reportajes que he publicado



#### **IV. Fuentes de investigación**

**P**ara la elaboración de esta investigación y tesis doctoral se han seguido y analizado diferentes y variadas fuentes de información. La literatura, en forma de estudios, artículos científicos y periodísticos o libros, ha dotado de ideas y contenido a esta tesis. La documentación técnica de los diferentes proyectos y procesos de automatización analizados ha sido la base para la exposición y la descripción de las nuevas tecnologías aplicadas al periodismo.

Episodios de podcast, boletines de información o conversaciones con especialistas sobre datos, automatización y el uso de la tecnología en la redacción también han contribuido a dar forma a esta tesis, dotándola de anécdotas y exposiciones más precisas sobre los elementos más complejos.

Debido a la temática que cubre esta tesis, cabe mencionar que gran parte de la información recabada pertenece a publicaciones de medios de comunicación americanos. Por tanto, existe una tendencia a reflejar los hábitos y técnicas del periodismo estadounidense.

Pese a ello, esta disertación trata de incluir voces y conocimiento sobre cómo se está integrando la automatización en las redacciones en otros países, en especial en España.

Por otro lado, el periodismo robot que se estudia en esta tesis es una disciplina relativamente nueva y novedosa en el sector del periodismo. Por ello, esta disertación contiene un elevado número de referencias a artículos periodísticos publicados en los últimos años debido a la limitada literatura que existe en el ámbito del periodismo y de comunicación la para describir algunos de los elementos analizados.

Respecto a las fuentes bibliográficas se han consultado un buen número de libros sobre periodismo, artículos científicos, revistas y publicaciones en línea. Como se ha mencionado previamente, y a pesar de que se ha intentado mostrar una visión global del estado de la cuestión estudiado en esta tesis, existe una tendencia a reflejar las prácticas del periodismo americano por ser el más avanzado en la cobertura y desarrollo de algoritmos, la automatización y el periodismo robot en general.

Además de fuentes bibliográficas, también se han utilizado testimonios de entrevistas que he realizado y conferencias a las que he asistido. Esta disertación ha sido completada con declaraciones, conversaciones y entrevistas realizadas durante los últimos años a profesionales y académicos en periodismo de datos, periodismo computacional o automatización. Estas entrevistas e intercambio de ideas me han ayudado a explorar nuevos caminos y han dado forma a la presente disertación.

En este sentido, se ha elaborado una hemeroteca de artículos periodísticos sobre los diferentes temas abarcados en esta disertación — como son la automatización de procesos, el análisis de datos, el uso de la inteligencia artificial o la investigación de algoritmos — con el fin de dotar de una estructura y un orden al análisis de contenidos de las piezas seleccionadas. De esta forma, y mediante un análisis estructurado de la información, se persigue un estudio objetivo de los contenidos informativos recopilados para la elaboración de esta tesis.

Estas piezas han servido para ejemplificar las diferentes ramificaciones y tendencias del periodismo robot estudiado en esta tesis.

En el análisis de esta nueva disciplina también está plasmada mi propia experiencia y los conocimientos que he adquirido tanto en instituciones académicas como en medios de comunicación en España, Reino Unido y Estados Unidos.

Toda mi carrera profesional ha estado ligada a contar historias con datos. Por tanto, durante este tiempo he experimentado muchas de las técnicas relatadas en esta tesis. Esta experiencia me ha sido de gran ayuda para entender el funcionamiento de estas técnicas y poder explicarlas con exactitud en el campo académico de la comunicación <sup>11</sup>.

---

<sup>11</sup> La experiencia propia y las observaciones personales han jugado un papel fundamental en el desarrollo de esta tesis. En primer lugar, mis estudios de Periodismo en la Universidad Complutense de Madrid (2009-2014) y mi estancia en *Loyola University of Chicago* (2013) despertaron mi interés por investigar nuevas formas de hacer periodismo. Mis estudios de posgrado, *MS in Data Journalism* en la Universidad de Columbia (2019), confirmaron mi interés y desarrollaron mis conocimientos por el periodismo de datos, la visualización y la automatización de procesos en el sector del periodismo. Por último, el MBA in International Management que cursé en la Universidad Internacional Menéndez Pelayo (2018) me ha permitido analizar el sector del periodismo desde un punto de vista de la empresa informativa.

Además, mi experiencia profesional en medios como *The Washington Post* (2019 - Presente), *El Confidencial* (2015-2018), *The Times of London* (2015), *Agencia EFE* (2014-2015) y *Radio Televisión Española* (2014) también han dado forma a esta tesis y los casos que se mencionan en ella.



## V. Hipótesis de partida

La presente tesis busca sentar las bases para analizar el uso de la tecnología en el periodismo y, por ende, de la automatización de procesos en el sector. La intersección de otras disciplinas como la ciencia de datos, la ingeniería informática o las matemáticas con el periodismo ha trazado nuevos caminos para la profesión y ha abierto la puerta a múltiples colaboraciones entre disciplinas y perfiles profesionales en los últimos años.

Las aplicaciones tecnológicas en el periodismo han cambiado sector. El periodismo de la última década no puede entenderse sin la automatización de procesos, el uso de algoritmos o la inteligencia artificial. Como consecuencia, esto ha afectado de forma radical a la estructura de las redacciones y al perfil de los profesionales de la información.

La tecnología y el periodismo robot han potenciado disciplinas como el periodismo de investigación o el periodismo visual. En la mayoría de los reportajes de investigación basados en datos y publicados por grandes medios internacionales, alguna o varias partes del proceso de producción fue creada de forma automática. Ya sea la recopilación de los datos en los que

se basa el artículo, el análisis o la visualización de estos, dichas historias cuentan con un elemento de automatización en su producción como se analiza en esta disertación.

La evolución del periodismo robot y la automatización de procesos tiene su reflejo en diferentes ramas del periodismo. La automatización se puede aplicar tanto para mejorar procesos en la redacción, como la lectura y procesamiento de bases de datos que alimentan las historias periodísticas o alertan al periodista de un suceso, como para investigar historias mediante el uso y desarrollo de algoritmos.

La automatización de procesos también optimiza la creación de historias visuales, gráficos y mapas que permiten crear nuevas dinámicas y experiencias para contar historias a través del lenguaje visual.

Las técnicas mencionadas en estas líneas implican el conocimiento de diferentes lenguajes de programación y el perfeccionamiento en su uso. Además del propio idioma hablado y escrito, los reporteros de hoy en día también saben comunicarse con los ordenadores y máquinas. Por poner algunos ejemplos, Javascript, R, Python o bash son algunos de los lenguajes más populares entre los profesionales de las secciones de Datos, Gráficos o Diseño de los principales medios internacionales.

El conocimiento de idiomas siempre es importante para un periodista, pero el conocimiento de idiomas tecnológicos se ha convertido en un elemento fundamental en las redacciones. El perfil del periodista, especialmente en Estados Unidos, ya no es sólo el de un profesional con un bagaje en humanidades o ciencias sociales. Cada vez más profesionales provienen de otras disciplinas como la ingeniería informática o la estadística.

¿Qué supone ello? Una ventaja competitiva a la hora de no sólo controlar la palabra escrita sino también los lenguajes que le van a permitir amplificar el impacto de sus técnicas y conocimientos periodísticos.

Tras la observación minuciosa de los cambios que se han producido en el sector del periodismo en los últimos años, esta tesis pretende explorar y demostrar las siguientes hipótesis:

- El periodismo robot ha abierto nuevos caminos para los medios de comunicación y el profesional de la información. La presencia de cada vez más algoritmos que tienen un efecto en la sociedad ha abierto nuevas oportunidades para investigar este tipo de cuestiones. Además, nuevos géneros periodísticos han surgido en torno a la expansión del periodismo robot.
- El sector del periodismo, en continua evolución, ha sufrido una de las transformaciones más veloces y significativas durante la última década con la incorporación de nuevas tecnologías. Esta transformación vertiginosa ha tenido un efecto en el negocio del periodismo y en la productividad de sus profesionales.
- El conocimiento de técnicas computacionales y lenguajes de programación es fundamental para aquellos periodistas que trabajan con datos. Hoy en día, para explicar problemas complejos, no sólo vale con manejar el arte de la palabra, también es necesario sobresalir tanto en el análisis de datos como en el lenguaje visual.
- La incorporación de técnicas como la automatización de procesos o el uso de algoritmos ha supuesto más beneficios que desafíos para los periodistas y el sector del periodismo.
- Estas tecnologías y su aplicación en el periodismo es todavía relativamente nueva. Por ello, todavía existen una serie de desafíos e interrogantes alrededor de la disciplina del periodismo robot.



# **CAPÍTULO I:**

## **Los inicios**



**E**l rol actual del periodista que trabaja con datos no es consecuencia de unos pocos años de absorción de la tecnología por el sector del periodismo. Por el contrario, es un proceso mucho más largo y complejo de más de medio siglo de incorporación paulatina de la tecnología al periodismo.

Los periodistas se han beneficiado del uso de los datos, y por ende de la automatización de procesos, desde que se popularizase el uso de los ordenadores en las redacciones de los medios de comunicación americanos a finales de los años 50.

Bien es cierto que, en la última década, el perfil tipo del periodista y su relación con la tecnología ha evolucionado a mayor velocidad que en el pasado. Hace unos años, el producto final de un reportero eran sus textos y su principal objetivo era escribir claro y conciso. Hoy en día, el sector del periodismo lo copan perfiles mixtos, multidisciplinarios y que pueden encargarse de diferentes tareas más allá de la redacción periodística. Un periodista ya no sólo escribe textos. Un periodista también se encarga de recopilar datos, crear herramientas de las que pueden beneficiarse otros compañeros o trabajar en proyectos de innovación para mejorar

el alcance de los medios de comunicación. El bagaje de un periodista actual puede provenir de lo ingeniería informática, la ciencia de datos, las matemáticas o cualquier otra disciplina alejada de las ciencias sociales.

Un día cualquiera en la vida de un periodista de datos, puede incluir la ejecución de diferentes procesos como la extracción de datos, el análisis y la visualización de los mismos, el diseño del artículo y la redacción de la historia, entre otros. Todos y cada uno de estos procesos requieren en la mayoría de los casos del conocimiento y uso de un lenguaje de programación, así como de la automatización para reducir tiempos y hacer posible el procesamiento de grandes bases de datos.

Tabla 1. Nuevas tareas del periodista como consecuencia de la absorción de diferentes procesos tecnológicos.

<b>Investigación</b>	Investigación de contenidos	Propuesta de temas periodísticos		
<b>Datos</b>	Extracción de datos	Limpieza de datos	Análisis de datos	Visualización de datos
<b>Diseño y visualización</b>	Diseño de páginas web	Diseño de gráficos y mapas		
<b>Redacción</b>	Redacción de textos	Integración de otros elementos		
<b>Desarrollo</b>	Creación y mantenimiento de APIs	Creación de herramientas y aplicaciones		

Fuente: Elaboración propia.

Este perfil multidisciplinar, al que se suele hacer referencia en la literatura como periodista de datos ha florecido en los últimos años en las redacciones de los medios de comunicación.

De esta forma, el periodismo de datos, el periodismo computacional o, en el caso de esta tesis, el periodismo robot, no son disciplinas recientes. Para conocer sus inicios y su evolución hay que remontarse a los años 50 y analizar uno de los primeros fenómenos en el periodismo estadounidense en el que unos profesionales de la información se sentaron frente a un ordenador para procesar y analizar datos.

Aunque la segunda mitad del siglo XX es el período en el que el uso y el análisis de los datos disponibles se introducen en el periodismo moderno, HOWARD argumenta que el uso de los datos en la comunicación es anterior <sup>12</sup>. De acuerdo con HOWARD, algunos de los precursores de estos métodos son los mercaderes italianos que durante el siglo XVII mantenían un registro del precio de las mercancías o el ingeniero y economista escocés, William Playfair (1759-1823), que desarrolló una serie de métodos gráficos para visualizar números, valores matemáticos y estadísticas <sup>13</sup>.

---

<sup>12</sup> Howard, A.B. (2014) *The Art and Science of Data-Driven Journalism*. Tow Center for Digital Journalism. Consultado a 16 de Enero de 2022, disponible en: <https://academiccommons.columbia.edu/doi/10.7916/D8Q531V1>

<sup>13</sup> Berkowitz, B. D. (2018). *Playfair: The True Story of the British Secret Agent who Changed how We See the World*. United States: George Mason University Press.

## 1.1. Un primer intento para predecir el resultado de una elección

Uno de los primeros hitos de los que se tiene constancia en la intersección del periodismo y la tecnología es el uso del ordenador QUES Remington Rand UNIVAC por la cadena de televisión estadounidense Columbia Broadcasting System (CBS). Mediante el uso de este ordenador, los periodistas de CBS pretendían predecir el resultado en las elecciones presidenciales de 1952 entre el candidato republicano, Dwight D. Eisenhower, y el candidato demócrata, Adlai Stevenson II <sup>14</sup>.

Esta iniciativa es uno de los primeros ejemplos donde se puede ver la relación y colaboración entre disciplinas y técnicas como el periodismo, los datos, las ciencias computacionales, el uso de los gráficos y el periodismo visual.

---

<sup>14</sup> Diakopoulos, N. (15 de abril de 2020). *The ethics of Predictive Journalism*. Columbia Journalism Review. Consultado a 5 de abril de 2021, disponible en: [https://www.cjr.org/tow\\_center/predictive-journalism-artificial-intelligence-ethics.php](https://www.cjr.org/tow_center/predictive-journalism-artificial-intelligence-ethics.php)

Figura 1. Artículo publicado el 15 de Octubre de 1952 en el diario Philadelphia Evening Bulletin sobre el uso del ordenador QUES Remington Rand UNIVAC por la cadena CBS para predecir los resultados de la elección presidencial de Estados Unidos de 1952. <sup>15</sup>

THE EVENING BULLETIN, PHILADELPHIA

### Network 'Drafts' Univac for Election Coverage



Univac, the electronic computer which will be used by the Columbia Broadcasting System in its television coverage of the November 4 elections, gets a test run under the supervision of (from left) Harold Sivoney, chief operator; J. Prespar Eckert, one of the machine's inventors, and CBS commentator Walter Cronkite, at the Eckert-Mauchly Computer Corp., 3747 Ridge av.

#### Stevenson

*Continued From First Page*

and ruthlessly on misconduct wherever I have found it.

"I even have a trained investigator in my office whose job it is to look out for it all the time."

#### Touches on Korea

And in reply to what he said was an implication by Eisenhower that "we could bring our men home from Korea soon if we could only train some South Koreans to take their place," Stevenson said at Spokane that this was one of the "strange things" in the campaign.

Stevenson said he had heard from Washington that more than half the divisions in the front lines in Korea are South Koreans, that we are steadily training the South Koreans to take over, but that this will take some time. If we were to pull back our troops and let the South Koreans take over, it would mean the fall of Korea and perhaps most of Asia, he said.

Stevenson said he "deplores" any suggestion that our men can come home any sooner than "our national safety permits." And he added: "There is no trick that can end the Korean war—and I am sure the general knows this full well."

#### Economy in Government

### CBS to Use Electronic Robot To Forecast Election Results

An electronic robot with a prodigious memory will go to work for a television audience November 4.

Officials of the Columbia Broadcasting System hope that the machine, known as Univac, will make it possible for them to forecast the final outcome of the elections on the basis of incomplete returns.

The \$500,000 electronic computer was made in Philadelphia by the Eckert-Mauchly division of Remington Rand, Inc., 3747 Ridge av.

#### Maze of Wires, Tubes

Univac is a maze of 50 miles of wire, 500,000 soldered joints and 5,000 vacuum tubes. It can add 2,000 numbers of 12 digit each in one second and can multiply and divide equally fast.

It should be just the thing, CBS thinks, for figuring out the election returns.

From now until election night Univac, short for universal automatic computer, will absorb figures on hourly-hourly election totals of 1944 and 1948 for each of the 48 states, plus detailed data on eight key states.

The machine will use these figures as comparisons when the first returns begin coming in on the night of November 4.

#### Should Provide Forecast

And in a few seconds, it is hoped, it will come up with a forecast of what the final 1952 totals will be from each state. Univac will take into account, while doing this, the variations in Democratic-Republican ratios that usually occur in later returns.

As Univac gets more complete material, later in the evening, the margin of error will be cut down.

The computer is controlled by an operator sitting at a central control unit that looks somewhat like the console for a pipe organ. The brain of the device is an insulated cabinet eight feet high, eight feet wide and 14 feet long.

CBS hopes that Univac will give fairly conclusive clues on the final results earlier than usual, with an added lift for the television audience which will see the monster in operation.

But, add the officials, it will be backed up by the traditional, old-fashioned method of tabulation—counting and reporting the election returns by hand.

Fuentes: Penn Libraries, University of Pennsylvania y Philadelphia Evening Bulletin.

<sup>15</sup> Disponible en Penn Libraries, University of Pennsylvania <https://www.library.upenn.edu/collections/online-exhibits/john-w-mauchly/univac-legacy-ibm/item=2>

Los ingenieros J. Presper Eckert y John Mauchly crearon durante los años 40 en la universidad de Pensilvania <sup>16</sup> un prototipo de lo que más tarde sería el ordenador *QUES Remington Rand UNIVAC*. La compañía estadounidense *Remington Rand* compró el ordenador y en 1952, después de haber vendido su primer ordenador a la oficina del Censo de Estados Unidos un año antes, se lo ofreció a la cadena de televisión *CBS News*. La propuesta para la venta del equipo fue predecir los resultados electorales de 1952 <sup>17</sup>.

Eckert y Mauchly trabajaron con su compañero de universidad Max Woodbury, matemático, para programar el algoritmo con el que pretendían predecir el resultado de las elecciones.

A las 20:30 horas de la costa este de Estados Unidos del 4 de noviembre de 1952, y después de introducir los datos manualmente en el ordenador, *Univac* predijo que Eisenhower ganaría las elecciones a Stevenson por 438 delegados electorales frente a 93.

Como las encuestas electorales habían predicho una victoria holgada del candidato demócrata y gobernador de Illinois, Adlai Stevenson, *CBS News* no anunció la predicción hasta una hora más tarde cuando Eisenhower comenzó a ganar varios estados y se posicionó como el favorito para vencer en las elecciones.

---

<sup>16</sup> Alfred, R. (4 de noviembre de 2010). *Nov. 4, 1952: Univac gets election right, but CBS balks*. *Wired*. Consultado a 5 de febrero de 2022, disponible en: <https://www.wired.com/2010/11/1104cbs-tv-univac-election/>

<sup>17</sup> Shedden, D. (6 de noviembre de 2014). *Today in media history: In 1952, a computer helped CBS predict the winner of the presidential election*. *Poynter*. Consultado a 5 de febrero de 2022, disponible en: <https://www.poynter.org/reporting-editing/2014/today-in-media-history-in-1952-a-univac-computer-helped-cbs-news-predict-the-winner-of-the-presidential-election/>

El resultado final fue de 442 votos electorales para Eisenhower y 89 para Stevenson. La predicción del ordenador tan sólo cometió un error del 1% <sup>18</sup>.

A pesar de que las comparaciones son odiosas con los métodos utilizados hoy en día, este hito se considera como el inicio del uso de los métodos de computación para la creación de informaciones en el sector del periodismo.

## 1.2. Periodismo de precisión

El periodista y profesor de periodismo estadounidense, Philip Meyer, profundizaría dos décadas más tarde en el uso de los datos y las técnicas de análisis por parte de los periodistas en su libro *Precision Journalism: Reporter's Introduction to Social Sciences* <sup>19</sup>. En dicho manual, Meyer propone la aplicación de métodos de otras ciencias para contar historias periodísticas.

Meyer propuso y desarrolló una serie de buenas prácticas para aplicar el análisis de datos en la disciplina periodística mediante el uso de la tecnología disponible en los años 60 y 70 en Estados Unidos. De acuerdo con Tejedor y Dader, Meyer defendió la aplicación de técnicas provenientes de disciplinas externas como la sociología, la estadística o la computación e el periodismo. Meyer fue uno de los precursores que vio el potencial que suponía la aplicación de otras disciplinas en el periodismo de investigación <sup>20</sup>.

---

<sup>18</sup> Goff L. (16 de febrero de 1999). *Univac predicts winner of 1952 election*. CNN. Consultado a 6 de febrero de 2022, disponible en: <http://www.cnn.com/TECH/computing/9904/30/1952.idg/>

<sup>19</sup> Meyer, P. (1979). *Precision Journalism: A Reporter's Introduction to Social Science Methods*. United Kingdom: Indiana University Press.

<sup>20</sup> Tejedor, L. y Dader, José Luis. (2011). *El patrimonio del César ante el Perro Guardián: un análisis de Periodismo de Precisión sobre la transparencia económica de los miembros del Gobierno en España*. En *Estudios sobre el Mensaje Periodístico*, volumen 17, número 2, pp. 431-455. Disponible en: <http://revistas.ucm.es/index.php/ESMP/article/view/38124/36876>

En sus propias palabras, Meyer definió el periodismo de precisión como la “aplicación de métodos científicos de investigación social y del comportamiento a la práctica del periodismo”<sup>21</sup>.

De acuerdo con Dader y Gómez Fernández, Meyer propone un nuevo método periodístico que “hará hincapié en la revisión técnica de la metodología científica empleada en cualquier tipo de cuantificación susceptible de trascendencia noticiosa”<sup>22</sup>.

Uno de los ejemplos más conocidos sobre los inicios de estas técnicas en el periodismo es la investigación del propio Meyer sobre el coste de los seguros escolares en el entonces condado de Dade en Florida (Estados Unidos). Galindo Arranz cita dicha investigación como uno de los primeros ejemplos de periodismo de precisión<sup>23</sup>, génesis del periodismo robot de hoy en día que estudia esta tesis.

El método seguido por Meyer en esta investigación tendrá un impacto fundamental en su percepción del periodismo de precisión futura. En 1959, Meyer reparó en una queja de una comunidad de Miami que le puso sobre la pista del alto coste de los seguros escolares contra los incendios y los huracanes.

Meyer indagó los posibles elementos que podrían contribuir a ese sobreprecio. Para investigar esta historia<sup>24</sup>, Meyer planteó tres simples hipótesis, propias del delito de cohecho y tráfico de influencias, como explica en su libro *Precision Journalism: Reporter's Introduction to Social Sciences*:

---

<sup>21</sup> Meyer, P. (1979). *Precision Journalism: A Reporter's Introduction to Social Science Methods*. United Kingdom: Indiana University Press, p. 2.

<sup>22</sup> Dader, J.L. y Gómez Fernández, P. (1993). *Periodismo de precisión: Una nueva metodología para transformar el periodismo*, en *Articles*, análisis 15, pp. 99-116.

<sup>23</sup> Galindo Arranz, F. (2004). *Propuesta de periodización histórica y evolución conceptual del Periodismo de Precisión*, Universidad de Santiago de Compostela. Disponible en [http://webs.ucm.es/info/emp/Numer\\_10/Sum/3-04.pdf](http://webs.ucm.es/info/emp/Numer_10/Sum/3-04.pdf)

<sup>24</sup> Meyer, P. (1979). *Precision Journalism: A Reporter's Introduction to Social Science Methods*. United Kingdom: Indiana University Press, p. 12.

- Los agentes de las aseguradoras fijan precios más caros para los seguros escolares que para otro tipo de seguros.
- Los miembros del consejo que rige las escuelas del condado reciben algo de valor a cambio de aceptar el elevado coste de los seguros escolares.
- Las personas cercanas a los miembros del consejo que rigen las escuelas, es decir familiares, amigos o conocidos, reciben algo de valor que comparten con los miembros del consejo a cambio de que estos aceptasen los elevados costes de los seguros.

Para demostrar sus hipótesis, Meyer utilizó datos de múltiples fuentes y directorios recabando información sobre la financiación de las campañas políticas y los miembros de la junta de directores de las escuelas del distrito <sup>25</sup>. Meyer trató de demostrar los tres planteamientos expuestos anteriormente analizando la información de la que disponía. Como describe Galindo Arranz, Meyer descubrió cruzando datos que las compañías de seguros financiaban las campañas de reelección de los miembros del consejo que dirigía las escuelas <sup>26</sup>.

---

<sup>25</sup> Ibid.

<sup>26</sup> Galindo Arranz, F. (2004). *Propuesta de periodización histórica y evolución conceptual del Periodismo de Precisión*, Universidad de Santiago de Compostela. Disponible en [http://webs.ucm.es/info/emp/Numer\\_10/Sum/3-04.pdf](http://webs.ucm.es/info/emp/Numer_10/Sum/3-04.pdf)



Un puñado de periodistas siguió la propuesta de periodismo de precisión de Meyer <sup>27</sup>. Las técnicas que utilizaron para investigar sus historias y realizar sus reportajes van desde el uso de encuestas y cuestionarios sistemáticos hasta el uso de ordenadores para analizar grandes volúmenes de datos. Estas técnicas siguen presentes hoy en día en la mayoría de las redacciones como esta tesis presenta y se han expandido mediante la introducción de técnicas más avanzadas como el uso de algoritmos, la aplicación de la inteligencia artificial o la incorporación de los procesos automatizados.

Entre los años 60 y 90, el número de artículos basados en el periodismo de precisión, que pasará a llamarse periodismo asistido por ordenador a comienzos de los 80, prolifera en los medios norteamericanos. El periodismo de precisión evoluciona gracias a las nuevas generaciones de periodistas que encontraron en los ordenadores un soporte para analizar datos y contar historias que años antes hubiese sido imposible investigar.

Durante estas tres décadas se publican algunas de las historias más reconocibles del periodismo de precisión, que más tarde se conocerá como periodismo asistido por ordenador <sup>28</sup>. Elliot Jaspin y Maria Miro Johnson publicaron en el diario Providence Sunday Journal una historia donde, tras cruzar y combinar varias bases de datos, identificaron una serie de conductores de autobuses escolares con multas de mala conducción y antecedentes penales <sup>29</sup>.

Bill Dedman publicó en 1988 una investigación titulada *The Color of Money* (“El color del dinero”) <sup>30</sup>, con la que ganó el premio Pulitzer por periodismo de investigación,

---

<sup>27</sup> Bounegru, L. y Grey, J. (Eds.). (2012). *The Data Journalism Handbook: Towards A Critical Data Practice*. Amsterdam University Press, p. 237. Disponible en: [10.1017/9789048542079](https://doi.org/10.1017/9789048542079)

<sup>28</sup> Cox, M. (2000). *The Development of Computer-Assisted Reporting*. Estudio presentado en el coloquio Newspaper Division, Association for Education in Journalism and Mass Communication, Southeast Colloquium el 17 de Marzo de 20000 en la University de Carolina del Norte, Chapel Hill.

<sup>29</sup> Jaspin, E. G. y Johnson, M.M. (1 de Marzo de 1987). *R.I., system fails to fully check driving records of bus applicants*, Providence Sunday Journal.

<sup>30</sup> La investigación *The Color of Money* publicada por *The Atlanta Journal* y *The Atlanta Constitution* puede consultarse en el portfolio de su autor, el periodista Bill Dedman. Disponible en [http://powerreporting.com/color/color\\_of\\_money.pdf](http://powerreporting.com/color/color_of_money.pdf)

sobre la discriminación en la concesión de préstamos e hipotecas en los barrios de clase media norteamericana. Para ejecutar esos métodos, los periodistas recurrieron a la ayuda de los primeros ordenadores presentes en las redacciones norteamericanas.<sup>31</sup>

### 1.3. Periodismo asistido por ordenador (CAR)

Será en los años 80 cuando los planteamientos del periodismo de precisión planteado por Meyer se extiendan a más redacciones. Es en esta década cuando se acuña el término de *computer-assisted reporting* o periodismo asistido por ordenador (CAR)<sup>32</sup> con la fundación en 1975 de la organización de profesionales, *Investigative Reporters and Editors* (IRE) y en 1989 del *National Institute for Computer-Assisted Reporting* (NICAR) en Estados Unidos<sup>33</sup>.

IRE se fundó por parte de un grupo pequeño de reporteros que trabajaban en diferentes medios de comunicación y querían colaborar y compartir pistas y buenas prácticas sobre el periodismo de investigación<sup>34</sup>.

NICAR, una iniciativa de IRE apoyada por la universidad *Missouri School of Journalism*, surge como una respuesta al creciente número de reporteros practicando las técnicas del periodismo de precisión<sup>35</sup> para realizar investigaciones. En los años 80 y 90, los ordenadores habían evolucionado<sup>36</sup> y se había formado una red de periodistas que trabajaban

---

<sup>31</sup> McGregor, S. (2013). *Car hits the mainstream*. Columbia Journalism Review. Consultado a 16 de enero de 2022, disponible en: [https://archives.cjr.org/data\\_points/computer\\_assisted\\_reporting.php](https://archives.cjr.org/data_points/computer_assisted_reporting.php)

<sup>32</sup> Coddington, M. (2014). *Clarifying Journalism's Quantitative Turn*, Digital Journalism.

<sup>33</sup> Fink, K. y Anderson, C.W. (2015) *Data Journalism in the United States*. Journalism Studies, 16 (4). pp. 467-481.

<sup>34</sup> IRE. (s.f.). *About ire*. Investigative Reporters and Editors. Consultado a 17 de enero de 2022, disponible en: <http://www.ire.org/about-ire/>

<sup>35</sup> Ibid.

<sup>36</sup> Cox, M. (2000). The Development of Computer-Assisted Reporting. Estudio presentado en el coloquio Newspaper Division, Association for Education in Journalism and Mass Communication, Southeast Colloquium el 17 de Marzo de 20000 en la University de Carolina del Norte, Chapel Hill.

con bases de datos en diferentes medios de comunicación y que se ayudaban mutuamente para desarrollar nuevas técnicas de análisis de datos. Mediante conversaciones telefónicas o visitas a otros compañeros en diferentes medios, los periodistas estadounidenses como Brant Houston, Stephen Doig o Sarah Cohen aprendieron la profesión de forma colaborativa <sup>37</sup> <sup>38</sup>.

Tanto IRE como NICAR pusieron en marcha dos conferencias anuales de periodistas que llevan su mismo nombre con el fin de sentar las bases del periodismo asistido por ordenador, más conocido como periodismo de datos en la actualidad, y de las múltiples disciplinas surgidas en los últimos años.

Además de organizar estas conferencias anuales, los periodistas miembros de IRE y NICAR han contribuido a la expansión y evolución del periodismo publicando manuales y libros con las mejores prácticas para el uso de datos en periodismo. La profesión ha evolucionado gracias a manuales más recientes sobre cómo combinar textos y números en reportajes <sup>39</sup>, guías para familiarizarse con las bases de datos <sup>40</sup> o libros prácticos para aplicar los principios básicos del periodismo asistido por ordenador en la redacción <sup>41</sup>, entre otros.

Por último, aunque el uso de los ordenadores y la tecnología es fundamental en esta disciplina, cabe recordar las palabras de Elliot Jaspin, uno de los pioneros del periodismo asistido por ordenador que recoge Houston <sup>42</sup>:

---

<sup>37</sup> Houston, B. (13 de diciembre de 2021). *The history of Data Journalism*. DataJournalism.com. Consultado a 1 de mayo de 2022, disponible en: <https://datajournalism.com/read/longreads/the-history-of-data-journalism>

<sup>38</sup> Kelly, T., Houston, B. y Doig, S. (13 de diciembre de 2021.). *Episode 42: Conversation with Brant Houston and Stephen Doig*. SoundCloud. Consultado a 1 de mayo de 2022, disponible en: <https://soundcloud.com/datajournalism/episode-42-conversation-with-brant-houston-stephen-doig>

<sup>39</sup> Cohen, S. y Bruzzese, L. (2001). *Numbers in the Newsroom: Using Math and Statistics in News*. United States: Investigative Reporters and Editors, Incorporated.

<sup>40</sup> Herzog, D. L. (2015). *Data Literacy: A User's Guide*. United States: SAGE Publications.

<sup>41</sup> Houston, B. (2018). *Data for Journalists: A Practical Guide for Computer-Assisted Reporting*. United Kingdom: Taylor y Francis.

<sup>42</sup> Ibid

“Los ordenadores no convierten a un mal reportero en un buen reportero. Los ordenadores hacen mejor a un buen reportero”.

#### **1.4. La expansión del periodismo de datos**

La llegada de Internet a finales del siglo pasado supuso un empujón para el periodismo asistido por ordenador <sup>43</sup>. Con el paso del tiempo y la evolución de la disciplina las nuevas generaciones de periodistas comenzaron a incorporar técnicas más avanzadas de otros campos como la estadística, la sociología o la ciencia de datos. Incluso, estas nuevas generaciones acuñaron un nuevo término: el periodismo de datos.

En las dos últimas décadas, el periodismo de datos se ha convertido en un fenómeno a nivel global e internacional. A día de hoy, buena parte de las redacciones del mundo tienen al menos un periodista de datos o un equipo de datos.

Además de la expansión internacional, el trabajo de los propios periodistas y medios ha contribuido a sustituir el término de periodismo asistido por ordenador por el periodismo de datos. La primera edición del libro *Data Journalism Handbook* <sup>44</sup>, publicada en 2012, o los primeros blogs de nicho sobre periodismo de datos como *The Guardian Data Blog* <sup>45</sup> <sup>46</sup> del

---

<sup>43</sup> Segnini, G. (2008). *Introducción al Periodismo Investigativo y al Periodismo Asistido por Computadora*, en Agencia de los Estados Unidos para el Desarrollo Internacional (USAID), Guatemala, p. 67.

<sup>44</sup> Bounegru, L. y Grey, J. (Eds.). (2012). *The Data Journalism Handbook: Towards A Critical Data Practice*. Amsterdam University Press, p. 237. Disponible en: [10.1017/9789048542079](https://doi.org/10.1017/9789048542079)

<sup>45</sup> Chaparro Domínguez, M.A. (2013). La evolución del periodismo de precisión: el blog de The Guardian sobre periodismo de datos, en Actas In Congreso Internacional de la Comunicación y Sociedad Digital, Universidad Internacional de La Rioja, Logroño, 18 y 19 de abril de 2013. Disponible en: <https://dialnet.unirioja.es/servlet/articulo?codigo=4247831>

<sup>46</sup> Guardian News and Media. (10 de marzo de 2009). *The Guardian Datablog*. The Guardian. Consultado a 1 de mayo de 2021, disponible en: <https://www.theguardian.com/news/datablog/2009/mar/10/blogpost1>

diario londinense *The Guardian* o *The Nerd Blog* del medio de investigación estadounidense ProPublica <sup>47</sup> contribuyeron a expandir los conocimientos y la disciplina entre los profesionales de la información.

A nivel conceptual, FLORES y SALINAS definen el periodismo de datos como una convergencia de disciplinas cuyo fin último es la investigación de historias:

“De ahí que el periodismo de datos consiste en usar herramientas estadísticas y de visualización para contar mejor las viejas historias y descubrir nuevas historias que contar. Es, en opinión de muchos, la nueva veta del periodismo de investigación. Pero, el desarrollo y puesta en práctica de este tipo de periodismo requiere del conocimiento de otras disciplinas cuyos aprendizajes deben darse transversalmente en los planes de estudios de periodismo. Los antecedentes de la transversalidad de conocimientos, no sólo de contenido y de tecnología (con Internet a la cabeza), sino del anclaje con otras disciplinas de las Ciencias Sociales, como la Estadística o la Sociología, viene dado por la opinión de diversos expertos y de tendencias de este nuevo ecosistema” <sup>48</sup>.

A día de hoy, la idea que desarrolló Meyer hace medio siglo, no ha variado demasiado. El objetivo final del periodismo de precisión, el periodismo asistido por ordenador o el periodismo de datos es el mismo: servirse de los mejores métodos y disciplinas para desarrollar historias periodísticas que tengan un impacto en la sociedad.

---

<sup>47</sup> *The nerd blog*. (s.f.). ProPublica. Consultado a 1 de mayo de 2021, disponible en: <https://www.propublica.org/nerds>

<sup>48</sup> Flores Vivar, J. y Salinas Aguilar, C. (2018). *Algorithms, applications and Big Data, new paradigms in the process of communication and teaching-learning of data journalism*. En *Revista de Comunicación*, Volumen 17, número 2. Disponible en: [https://revistadecomunicacion.com/en/articulos/2018\\_2/12\\_text\\_Art.html#:~:text=De%20ah%C3%AD%20que%20el%20periodismo,veta%20del%20periodismo%20de%20investigaci%C3%B3n](https://revistadecomunicacion.com/en/articulos/2018_2/12_text_Art.html#:~:text=De%20ah%C3%AD%20que%20el%20periodismo,veta%20del%20periodismo%20de%20investigaci%C3%B3n).



# **Capítulo II: Periodismo robot**



**E**n los últimos años, la tecnología ha revolucionado el sector del periodismo tanto en su estructura como en sus contenidos. Estos cambios han afectado y se han hecho patentes también en el periodismo de datos. De dicha especialización, han surgido múltiples ramas, especializaciones y conceptos en el transcurso de apenas una década.

Estos términos derivados de disciplinas como el periodismo asistido por ordenador o el periodismo de datos presentan unos elementos en común en su fundación como son el uso de la computación y los datos. Cada uno de estos términos y conceptualizaciones pretende resaltar las características y técnica únicas o más relevante de las nuevas disciplinas.

Diakopoulos define computational journalism o periodismo computacional como la aplicación de técnicas y razonamientos de la computación en actividades periodísticas como la recolección de información, la estructuración de datos, así como la comunicación, presentación y diseminación de esa información. La precisión y la veracidad, propias de cualquier pieza periodística, son también principios fundamentales del periodismo computacional <sup>49</sup>.

---

<sup>49</sup> Diakopoulos, N. (2010) *A Functional Roadmap for Innovation in Computational Journalism*. School of Communication and Information, Rutgers University. Disponible en: <http://www.nickDiakopoulos.com/2011/04/22/a-functional-roadmap-for-innovation-in-computational-journalism/>

Coddington, por su parte, propone que el periodismo asistido por ordenador, el periodismo de datos y el periodismo computacional forman parte de una categoría superior como es el periodismo cuantitativo o quantitative journalism <sup>50</sup>. Una disciplina periodística que apuesta por la aplicación de mediciones objetivas y de análisis estadístico a la labor del periodismo.

El medio estadounidense FiveThirtyEight, especializado en el análisis político, económico y deportivo desde un punto de vista estadístico, es el medio de referencia en lo que a métodos cuantitativos se refiere. Sus periodistas y editores han desarrollado incluso guías para la correcta aplicación del periodismo cuantitativo al periodismo <sup>51</sup>.

Coddington además ofrece una definición alternativa de periodismo computacional, estableciendo una distinción clara entre el uso general de datos en piezas informativas y la manipulación y el análisis de esos datos por parte del periodista mediante de métodos computacionales <sup>52</sup>:

“I define computational journalism here as a strand of technologically oriented journalism centered on the application of computing and computational thinking to the practices of information gathering, sense-making, and information presentation, rather than the journalistic use of data or social science methods more generally”.

---

<sup>50</sup> Coddington, Mark (2015) *Clarifying Journalism's Quantitative Turn*, Digital Journalism, 3:3, 331-348, DOI: [10.1080/21670811.2014.976400](https://doi.org/10.1080/21670811.2014.976400)

<sup>51</sup> La presentación *Quantitative Editing, a guide* de la editora cuantitativa de *FiveThirtyEight*, Laura Bronner, detalla algunos de los procesos que siguen los editores y periodistas de la publicación estadounidense especializada en análisis político, económico y deportivo. Disponible en la página web de la periodista e investigadora: <https://www.laurabronner.com/quant-editing>

<sup>52</sup> Coddington, M. (2015) *Clarifying Journalism's Quantitative Turn*, Digital Journalism, 3:3, 331-348, DOI: [10.1080/21670811.2014.976400](https://doi.org/10.1080/21670811.2014.976400)

Por otro lado, tampoco existe un consenso claro en la literatura sobre las diferencias y similitudes entre periodismo automatizado, periodismo algorítmico o periodismo robot. De hecho, son términos que suelen utilizarse indistintamente a la hora de referirse a las tareas de automatización y aplicación de algoritmos que se llevan a cabo en una redacción de un medio de comunicación.

Para sentar un marco de estudio, se presentan a continuación algunas de las acepciones más aceptadas en la literatura para estas disciplinas.

Como se expone en el Capítulo IV. Automatización en el periodismo de investigación no existe consenso a la hora de definir el periodismo algorítmico. Kotenidis y Veglis proponen una visión amplia de esta disciplina. Los autores apuntan hacia la corriente que destaca que el periodismo algorítmico es una evolución del periodismo de datos, enfocado en el análisis de algoritmos y en el desarrollo de capacidades de procesamiento del software más moderno aplicado al periodismo <sup>53</sup>.

Por otro lado, Graefe define al periodismo robot como la disciplina que permite producir y generar informaciones basadas en datos limpios, estructurados y fiables <sup>54</sup>.

Por su parte, Lokot y Diakopoulos definen el término periodismo robot como la automatización de las tareas periodísticas relacionadas con la redacción periodística, la identificación de contenidos noticiosos o el análisis de los datos que alimenta las informaciones. <sup>55</sup>.

---

<sup>53</sup> Kotenidis, E. y Veglis, A. (2021). *Algorithmic Journalism—Current Applications and Future Perspectives*, *Journalism and Media.*, 2021; 2(2):244-257. Disponible en: <https://doi.org/10.3390/journalmedia2020014>

<sup>54</sup> Graefe, A. (2016,). *Guide to automated journalism*. Columbia Journalism Review. Consultado a 4 de septiembre de 2021, disponible en: [https://www.cjr.org/tow\\_center\\_reports/guide\\_to\\_automated\\_journalism.php](https://www.cjr.org/tow_center_reports/guide_to_automated_journalism.php)

<sup>55</sup> Ibid.

## **2.1. Periodismo robot como término global**

Por estos motivos y falta de consenso, esta tesis propone y presenta un marco más amplio en torno a los procesos de automatización en el periodismo y en la generación de noticias. De esta forma esta disertación no sólo se centra en generación de textos por parte de un robot, definición tradicional de la disciplina a la que en los últimos años se ha denominado periodismo robot.

El marco de esta tesis es mucho más amplio e incluye y analiza cualquier proceso de automatización, hasta el más nimio, ejecutado en el proceso de producción de una noticia. Es decir, esos procesos pueden no estar relacionados con la generación del núcleo más importante de la noticia: el texto o la palabra. Estos procesos incluyen la extracción, actualización automática o el análisis de los datos que alimentan esa noticia, la producción de los gráficos que ilustran esa noticia o la creación de avisos automáticos relacionados con esa noticia, entre otros.

Esta disertación pretende analizar cómo la automatización de procesos en el periodismo, que se ha dado de manera vertiginosa en los últimos años, ha cambiado el paradigma y el rumbo del periodismo. En la actualidad, los grandes medios apuestan por una mayor integración de estos procesos en sus redacciones, así como un perfil mucho más variado de los profesionales que trabajan en ellas.

Por ello, y ante la intención de adecuar la discusión al paradigma actual periodístico, esta tesis amplía la definición del periodismo robot a cualquier proceso de automatización, desde el más simple al más complejo, durante el proceso de producción, presentación y publicación de una noticia. Al referirme a noticia, pretendo referirme a todos los elementos que componen un artículo de la era digital. Es decir, además del texto, audio o vídeo, se engloban otros elementos como el análisis de datos, la visualización, el diseño de la página o la interacción con el lector, en caso de que apliquen.

Como se ha visto en los artículos académicos revisados, el periodismo robot no cuenta con una definición cerrada debido a los múltiples usos que se le ha dado en la literatura. En cambio, todas estas conceptualizaciones sí engloban la principal característica que se analiza y estudia en esta disertación: la automatización de contenidos.

Por esta razón, esta tesis toma prestado el término periodismo robot al considerar que engloba todas las tareas tecnológicas, los procesos de automatización y el uso de algoritmos que se llevan a cabo en una redacción de un medio de comunicación actual. De la misma forma, esta tesis evita el uso de otros términos para referirse a estos procesos por ser su ámbito y mira mucho más estrecho o centrado en tareas muy específicas.

Por poner un ejemplo, el concepto de periodismo de datos, centrado en el análisis de los datos para contar historias, no abarca todos los matices y tareas que se abarcan en esta disertación. Por su parte, el periodismo algorítmico se entiende como una especialización más de nicho por centrarse primordialmente en el estudio o desarrollo de algoritmos.

## **2.2. Retrato robot de los nuevos profesionales de la información**

Una vez fijado el marco de estudio, cabe examinar qué tipo de cualidades o habilidades presenta el periodista actual que se desenvuelve y trabaja en la disciplina del periodismo robot. Según García-Alonso, la realidad mediática y de producción de contenidos a la que se enfrentan los periodistas actuales ha cambiado en los últimos años.

“Es preciso formar al nuevo periodista para poder acometer la situación diferente a la que se enfrenta: (...) hay que añadir un necesario enriquecimiento de la información, más trabajada, enriquecida con enlaces o links que permitan al lector aumentar su acceso hasta donde satisfaga su deseo”<sup>56</sup>.

---

<sup>56</sup> García-Alonso, P. (2012). “Empresa Informativa en 2012. Nuevo modelo de negocio”, pp. 1-9, Capítulo en Actas IV Congreso Internacional de Ciberperiodismo y Web 2.0.pdf (8.086Mb) 2012, Ed. Servicio editorial de Universidad del País Vasco (UPV/EHU), p.4. Disponible en: <https://dialnet.unirioja.es/servlet/articulo?codigo=7739731>

En este sentido, ¿cuáles son los conocimientos que debe poseer el periodista de datos moderno? ¿Qué herramientas forman parte de su trabajo diario?

Los periodistas especializados en periodismo robot y automatización, que aplican esas técnicas a la investigación de historias, el análisis y la visualización de datos, tienen en su mayoría un perfil multidisciplinar. Es decir, además de sus habilidades periodísticas, poseen conocimientos de otras disciplinas como la ingeniería informática, la estadística o la ciencia de datos.

Estos periodistas pueden trabajar en solitario o en equipos multidisciplinarios. Por lo tanto, sus conocimientos de estas disciplinas pueden ser básicos para poder colaborar con profesionales de otros equipos o amplios para trabajar individualmente.

La organización sin ánimo de lucro Data Visualization Society publica cada año los resultados de su Encuesta sobre el Estado de la Industria <sup>57</sup>.

La encuesta, completada por más de 2.100 profesionales de la información y del diseño de todo el mundo, es relevante ya que expone sus condiciones laborales y su forma de trabajar. Además, la encuesta de 2021 incide en cómo ha impactado la COVID-19 en sus trayectorias profesionales.

De acuerdo con los resultados de la última encuesta de 2021, la herramienta más popular y más utilizada por los profesionales de la información es el programa de hojas de cálculo Microsoft Excel.

Por otro lado, como muestra el estudio, al menos uno de cada cuatro entrevistados conoce y utiliza lenguajes de programación como JavaScript, y su librería de manipulación y

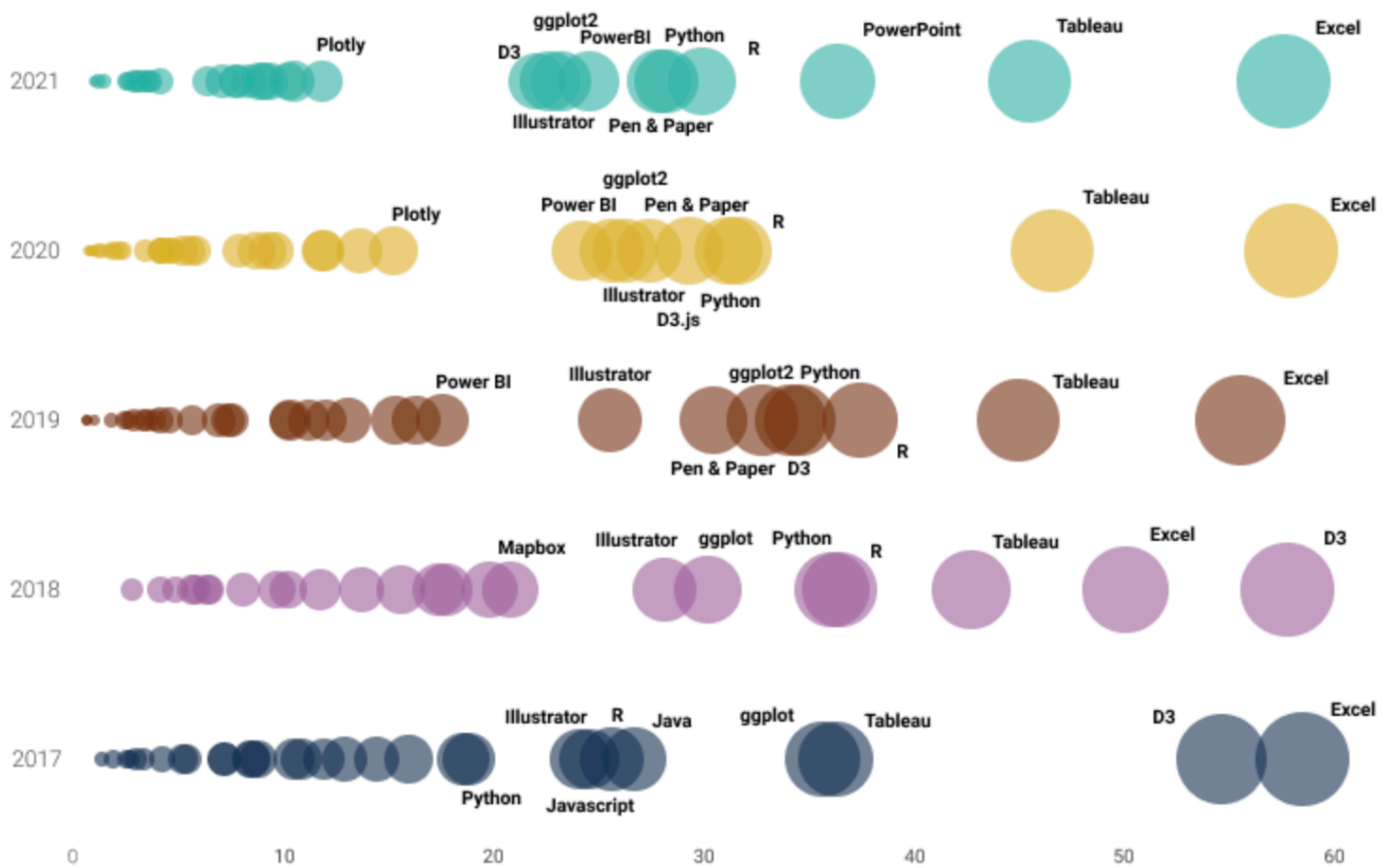
---

<sup>57</sup> Data Visualization Society. (2021) *State of the industry survey — 2021 report*. Data Visualization Society. Disponible en el portal: <https://www.datavisualizationsociety.org/report-2021> y en un formato .pdf en: [https://static1.squarespace.com/static/5c6055b5fb18206d45d6b27e/t/61f0573657c6471e4d53dc58/1643140919414/DVS\\_State+of+the+Industry+2021+Report.pdf](https://static1.squarespace.com/static/5c6055b5fb18206d45d6b27e/t/61f0573657c6471e4d53dc58/1643140919414/DVS_State+of+the+Industry+2021+Report.pdf)

visualización de datos d3.js, R, y su librería de visualización de datos ggplot2, o Python para el análisis de datos en su día a día laboral.

En el Capítulo VI. Automatización en el periodismo visual se profundiza en el uso y aplicación de estos lenguajes por parte de los profesionales de la información.

Figura 3. Gráfico realizado por Data Visualization Society con los resultados de su encuesta a la pregunta sobre qué tecnología o lenguaje de programación es el más usado por los profesionales de la información.



Fuente: *Data Visualization Society, State of the industry survey — 2021 report.*

Otras iniciativas en las que se pueden pulsar las tendencias y hábitos más comunes entre los profesionales de la información que trabajan con datos son la conferencias IRE y NICAR <sup>58</sup>.

Los contenidos de la conferencia NICAR son de especial interés en este sentido ya que es la conferencia por excelencia donde los profesionales de la información comparten las últimas tendencias para trabajar, analizar y visualizar datos desde el punto de vista del periodismo.

Las hojas de cálculo o Excel son también los conocimientos más demandados en ambas conferencias. Sin embargo, el programa de la conferencia de los últimos años ha evolucionado e incluye un creciente número talleres sobre tecnologías y lenguajes más avanzados como JavaScript, Python, R o SQL.

En conclusión, el periodista debe conocer diferentes lenguajes de programación y herramientas tecnológicas para poder trabajar en el entorno del periodismo robot. Los conocimientos multidisciplinarios de ámbitos como la ingeniería informática, la estadística y el diseño, entre otras, le permitirán ejecutar las tareas propias del periodismo robot.

---

<sup>58</sup> La información relativa a todas las conferencias IRE y NICAR puede consultarse en el portal web de IRE en la siguiente dirección web: <https://www.ire.org/training/conferences/> Los programas de las conferencias han sido analizados para determinar los talleres más repetidos a lo largo de los últimos años.



# **CAPÍTULO III:**

## **Automatización en la redacción**



**L**a introducción de procesos tecnológicos, computacionales y estadísticos complejos ha transformado las dinámicas y los flujos en las redacciones. Entre los principales cambios, destacan la creación nuevas oportunidades, proyectos innovadores y roles multidisciplinares para los profesionales de la información.

La automatización de procesos en la redacción consiste en automatizar o en derivar la ejecución de tareas repetitivas a un ordenador o robot, liberando a los periodistas de la realización periódica de esas tareas y procesos.

En el día a día en una redacción existen múltiples procesos que pueden realizarse de forma automática. Todos estos procesos, desde los más nimios, como descargar una base de datos de forma recurrente, hasta los más complejos, como escribir una noticia o analizar y visualizar unas elecciones, suponen, si bien ejecutados, una ayuda para el reportero.

### 3.1. Beneficios de la automatización

En primer lugar, la automatización de tareas ayuda a liberar de trabajo adicional al reportero, permitiendo que invierta su tiempo en la investigación o cobertura de temas clave para el medio de comunicación y los lectores.

De acuerdo con este argumento, los algoritmos partícipes en los procesos de automatización pueden, entre otras funciones, analizar datos, identificar historias interesantes para el lector y ofrecer un primer borrador del artículo al reportero. Una vez ejecutado este proceso, el periodista toma la iniciativa enriqueciendo la historia con un análisis en profundidad, entrevistando a personajes relevantes para la historia y recabando más datos e informaciones.

Graefe destaca que no es sólo una cuestión de ganar tiempo y liberar al reportero de algunas tareas, también supone una ventaja competitiva para la empresa informativa. El autor expone que en el actual mercado mediático la automatización es una necesidad debido a la cantidad de puertas que abre al reportero para trabajar en nuevas historias y a las oportunidades que ofrece al medio de comunicación para innovar en la creación de contenidos y mejorar los procesos internos en una redacción <sup>59</sup>.

Además, Graefe cita en su estudio a Gina Chua, editora ejecutiva para operaciones editoriales, datos e innovación en Thomson Reuters en el momento de publicación del estudio. Chua afirma que en con la actual composición y evolución del mercado de medios de comunicaron “no es posible competir sin automatizas tareas” <sup>60</sup>.

---

<sup>59</sup> Graefe, A. (2016,). Guide to automated journalism. Columbia Journalism Review. Consultado a 4 de septiembre de 2021, disponible en: [https://www.cjr.org/tow\\_center\\_reports/guide\\_to\\_automated\\_journalism.php](https://www.cjr.org/tow_center_reports/guide_to_automated_journalism.php)

<sup>60</sup> Ibid.

## 3.2. Riesgos de la automatización

Automatizar procesos, en cambio, sólo es beneficioso para el sector periodismo cuando se ejecuta bien. El diseño y supervisión editorial de los algoritmos es fundamental para que los resultados finales no contengan errores o su calidad sea inferior al producto que elaboraría un reportero sin la asistencia de la automatización.

Los procesos y algoritmos que no han sido diseñados correctamente o que no gozan de la supervisión adecuada también pueden tener un impacto en el producto final de la tarea de automatización, ya sea éste un artículo periodístico o una iniciativa para identificar nuevos temas o nuevos lectores.

El abuso o el mal uso de los algoritmos de automatización en la redacción puede derivar en burbujas de filtro en las que los algoritmos sugieren informaciones afines a la ideología de los lectores en lugar de basar el criterio de identificación y selección de noticias en la relevancia de las informaciones <sup>61</sup>.

La calidad de las informaciones y el lenguaje utilizado en ellas también puede verse afectado si el periodista no toma parte en el proceso de creación de estas informaciones o el algoritmo carece de recursos y opciones para elaborar un mejor producto periodístico <sup>62</sup>.

De igual modo que una noticia o un reportaje puede carecer de la calidad adecuada tras la intervención de un algoritmo, la automatización del procesamiento y el análisis de datos para la identificación y producción de historias puede conducir a conclusiones equivocadas o contener errores. Por ello, la participación del profesional de la información en este tipo de procesos es fundamental para fiscalizar la calidad de los resultados y mejorar la automatización de procesos en una redacción.

---

<sup>61</sup> Ibid.

<sup>62</sup> VV.AA. (2019). *News Automation The rewards, risks and realities of 'machine journalism'*. WAN-IFRA. Disponible en: [http://immersivautomation.com/wp-content/uploads/2019/06/WAN-IFRA\\_News\\_Automation-FINAL.pdf](http://immersivautomation.com/wp-content/uploads/2019/06/WAN-IFRA_News_Automation-FINAL.pdf)

### 3.3. La cobertura de elecciones

Siete décadas después de que el ordenador Remington Rand UNIVAC predijese para la cadena estadounidense CBS que el candidato republicano, Dwight D. Eisenhower, ganaría las elecciones de 1952, la cobertura de comicios electorales es todavía uno de los principales motores de las iniciativas de automatización en la redacción.

Durante este tiempo, los medios de comunicación y sus profesionales han evolucionado en el uso de la tecnología. Como indica García-Alonso, los medios deben abrirse a nuevos temas y mercados, aprovechando el acceso online, con nuevos contenidos <sup>63</sup>. Por ello, en la actualidad, los medios se sirven de la automatización de procesos para ofrecer nuevos contenidos.

De hecho, la automatización de procesos permite ofrecer el resultado del recuento electoral en vivo. Además, nuevos métodos estadísticos se aplican para predecir los posibles escenarios y resultados antes y durante una elección, dejando atrás las limitaciones del ordenador Remington Rand UNIVAC.

La cobertura de las elecciones ha jugado y juega un papel fundamental en la introducción de la tecnología en el sector del periodismo y en las redacciones. A diferencia del experimento llevado a cabo por CBS, donde tres profesionales de la información trataron de predecir el resultado de las elecciones de 1952, la actual apuesta de los medios de comunicación por la cobertura electoral involucra a múltiples equipos multidisciplinares con diferentes roles en una redacción.

En la última década, los intentos por predecir u ofrecer contexto antes y durante el recuento de votos de una elección se han intensificado.

---

<sup>63</sup> García-Alonso, P. (2018). “Nuevo modelo de negocio para la empresa informativa”, pp. 81-94, Capítulo en: “Perfiles actuales en la información y en los informadores”, Ed. Tecnos-Grupo Anaya, Madrid 2018, 1.ª edición, 2017, 497 págs. [https://repository.udem.edu.co/bitstream/handle/11407/5261/Perfiles\\_actuales\\_informacion\\_in...](https://repository.udem.edu.co/bitstream/handle/11407/5261/Perfiles_actuales_informacion_in...)

*The New York Times* desarrolló y publicó unas agujas para medir qué candidato está más cerca de ganar una elección durante el recuento del voto <sup>64</sup>. Gracias al procesamiento y análisis automático de los datos, previamente programado por los periodistas de *The New York Times*, es posible aplicar una serie de métodos estadísticos y visualizar las probabilidades de ganar la elección cada candidato.

Por su parte, *The Washington Post* ha apostado por visualizar la incertidumbre y los posibles escenarios de una elección durante el recuento del voto <sup>65 66 67</sup>.

---

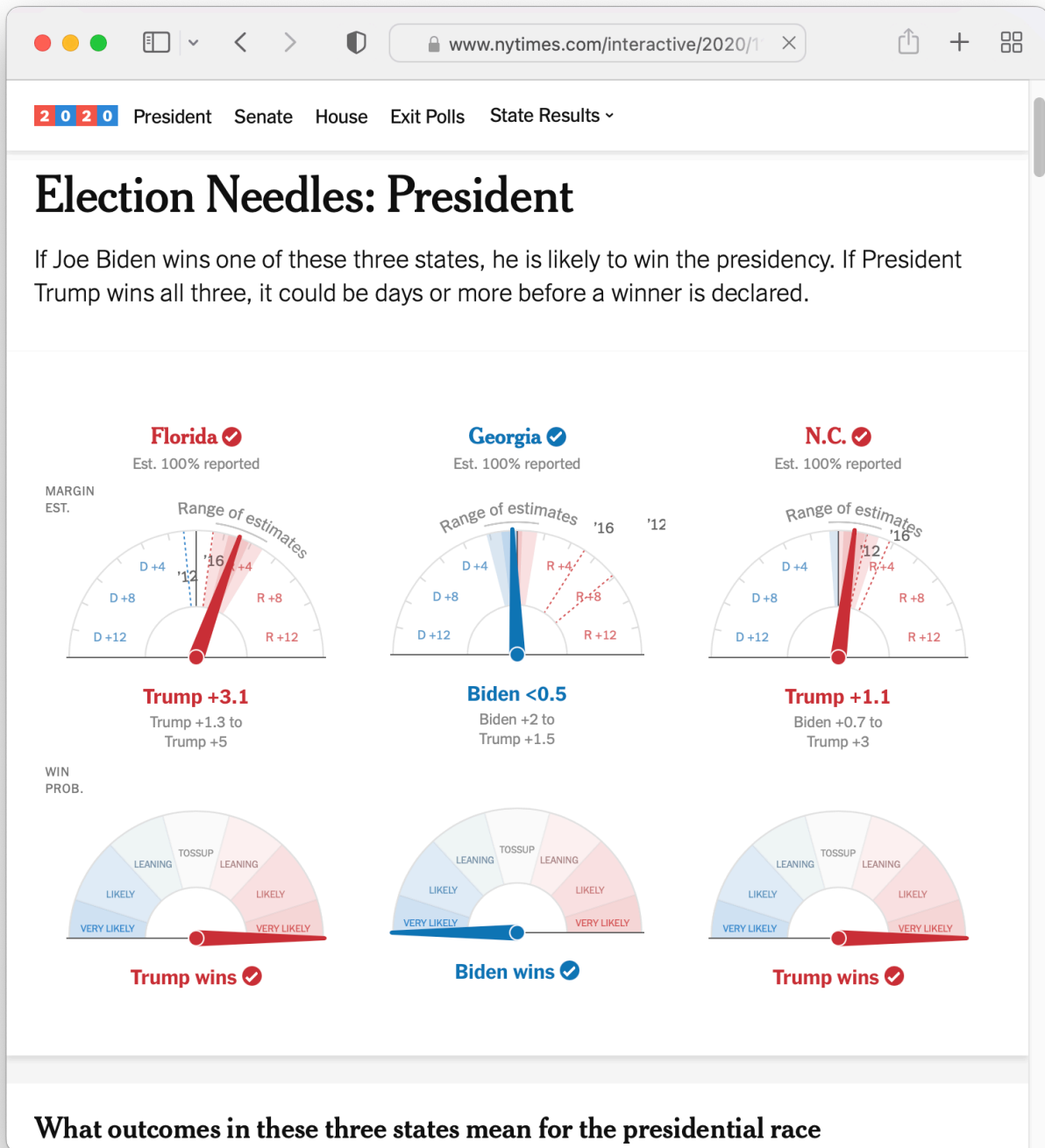
<sup>64</sup> The New York Times. (3 de noviembre de 2020). *Election needles: President*. The New York Times. Consultado a 10 de septiembre de 2021, disponible en: <https://www.nytimes.com/interactive/2020/11/03/us/elections/forecast-president.html>

<sup>65</sup> Bronner, L. (2 de noviembre de 2021). *How The Washington Post will model possible outcomes in the Virginia governor's race*. The Washington Post. Consultado a 11 de enero de 2022, disponible en: <https://www.washingtonpost.com/elections/2021/11/02/election-model-explained/>

<sup>66</sup> Cherian, J. y Bronner, L. (2020) *How The Washington Post Estimates Outstanding Votes for the 2020 Presidential Election*. Washington Post Engineering. Disponible en: [https://s3.us-east-1.amazonaws.com/elex-models-prod/2020-general/write-up/election\\_model\\_writeup.pdf](https://s3.us-east-1.amazonaws.com/elex-models-prod/2020-general/write-up/election_model_writeup.pdf)

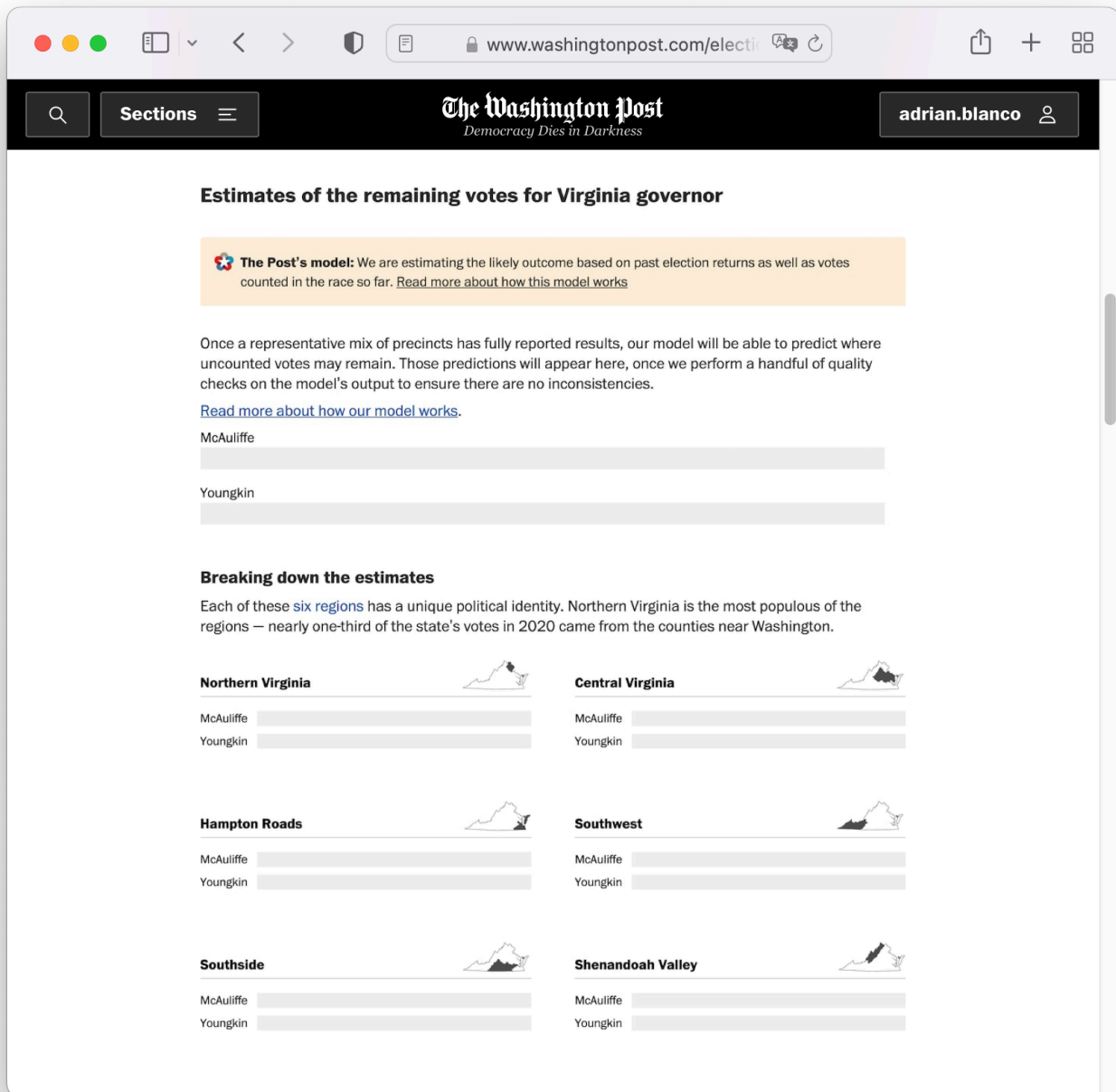
<sup>67</sup> Blanco, A. y Galocha, A. (24 de mayo de 2022). *How election modeling can help us understand who might win*. The Washington Post. Consultado a 24 de mayo de 2022, disponible en: <https://www.washingtonpost.com/elections/interactive/2022/how-election-models-work/>

Figura 4. Agujas diseñadas y utilizadas por el diario estadounidense *The New York Times* para ofrecer contexto y predecir los resultados de las elecciones presidenciales estadounidenses de 2020.



Fuente: *The New York Times*.

Figura 5. Apuesta visual de *The Washington Post* para ofrecer contexto y mostrar los diferentes escenarios posibles durante el recuento de votos en una elección. En este caso se muestran las elecciones a gobernador del estado de Virginia en 2021.



Fuente: *The Washington Post*.

Ambas iniciativas, prácticamente automatizadas en su totalidad por la labor de los profesionales de la información, suponen diferentes modos de despejar la incertidumbre de una elección e informar al lector. Pese a que las coberturas electorales por parte de los medios se han actualizado y son mucho más complejas que en los años 50, el objetivo sigue siendo el mismo: ofrecer información novedosa y valiosa al lector.

La cobertura de elecciones es además el evento mediático que más han contribuido a la participación de perfiles profesionales multidisciplinares en su desarrollo y a tender puentes entre equipos de periodistas, diseñadores e ingenieros para automatizar procesos en una redacción. Es por esta razón que se incluyen en esta sección como ejemplo de colaboración e intersección de diferentes conocimientos y disciplinas en una redacción periodística.

### **3.4. Automatización en forma de robots**

Los robots, más conocidos como bots, son otro de los ejemplos de cómo la automatización cada vez juega un papel más importante en el periodismo. Un bot una aplicación de software programada por una persona que ejecuta tareas estructuradas y repetitivas de forma automatizada.

Los bots son además un elemento muy flexible para los profesionales de la información debido a los múltiples roles que pueden ejercer y las diferentes funciones para las que pueden ser útiles en una redacción.

Los bots programados por un periodista o profesional de la información cuentan con objetivos muy diferentes. Tras la selección y el análisis de más de una treintena de bots desarrollados en el ámbito del periodismo y la comunicación <sup>68</sup>se propone la siguiente

---

<sup>68</sup> Consultar “Anexo IV. Selección de bots con fines periodísticos analizados en esta tesis” para ver una relación detallada de los bots seleccionados y analizados en esta disertación con el objetivo de crear una propuesta de taxonomía de estudio de los robots desarrollados en el ámbito de la comunicación.

taxonomía para su estudio que posteriormente se completa con una serie de ejemplos desarrollados por los profesionales de la información.

En primer lugar, los bots pueden tener una presencia pública, su objetivo es informar al lector, o interna, han sido desarrollados como una herramienta para la redacción. En función de la misión y las características de cada bot, la clasificación se expande hasta cubrir las principales funciones que tienen estas máquinas automatizadas en un medio de comunicación.

Figura 6. Clasificación propuesta para el análisis de bots en el ámbito del periodismo.

- **Bots internos**

- Bots para la manipulación de datos

- Extracción de datos
- Análisis de datos

- Bots para la ejecución de tareas

- **Bots públicos**

- Bots conversacionales

- Bots como herramienta informativa

- Bots en redes sociales

Fuente: Elaboración propia.

Por último, como se explica en la próxima sección, los bots públicos pueden ser parte del propio medio en forma de pieza periodística o tener una presencia social en redes como Twitter, Facebook Messenger o Telegram.

### **3.4.1 Bots de uso interno**

#### **3.4.1.1. Bots para la manipulación de datos**

En primer lugar, los profesionales programan archivos de código y secuencias de comandos para analizar datos que se publican de forma recurrente. Estos pequeños robots son muy útiles para analizar publicaciones de datos recurrentes y estacionales como las cifras del paro o desempleo publicadas por las diferentes oficinas estadísticas de cada país.

Los periodistas también hacen uso de los bots y la automatización para analizar conjuntos de datos más complejos. La creación de estos bots de análisis, que pueden ser reutilizados con algunos para otro tipo de datos o para futuras actualizaciones de los mismos datos, son muy útiles ya que reducen el tiempo en extraer conclusiones de los datos considerablemente.

En ocasiones, una vez analizados los datos, los periodistas liberan el código utilizado para el análisis con el fin de que pueda ser comprobado y auditado por otros colegas de profesión o por el público en general. A esta iniciativa en pos de la transparencia por compartir y liberar código se le ha denominado open source journalism <sup>69</sup>. Esta resolución, en cambio, todavía está lejos de ser una realidad global en todos los medios de comunicación por el tiempo adicional que requiere liberar los materiales y anonimizar los datos en caso de que sea necesario.

---

<sup>69</sup> Tracy, M. (1 de diciembre de 2019). *These reporters rely on public data, rather than secret sources*. The New York Times. Consultado a 12 de septiembre de 2021, disponible en: <https://www.nytimes.com/2019/12/01/business/media/open-source-journalism-bellingcat.html>

La herramienta más común para liberar el código es GitHub, una plataforma en la nube que permite compartir código y colaborar en su escritura . Un ejemplo de este tipo de prácticas es el código y los datos liberados por ProPublica para su investigación Machine Bias <sup>70</sup> que se examina más adelante en el Capítulo IV. Automatización en el periodismo de investigación.

Por otro lado, el data scraping o extracción de datos de la web es un proceso para obtener datos en exclusiva que no están disponibles mediante una descarga pública. El scraping de datos está basado en un script o secuencia de comandos que accede a la web que contiene la información deseada, identifica los datos y los descarga de forma automática <sup>71</sup>.

Los periodistas programan este tipo de bots o scripts para acceder a información que no está disponible de forma pública y que, por tanto, no puede obtenerse de otra forma.

Pese a que parece una técnica sencilla en el papel, el data scraping de datos es una técnica compleja por el número de problemas que el periodista puede encontrarse al acceder a una web de forma automática. Entre los problemas a resolver más comunes en la extracción de datos se encuentran <sup>72</sup>:

- Los archivos HTML — el esqueleto de las páginas web — mal formateados o con una estructura muy pobre provocan que el periodista tenga que buscar soluciones y prever diferentes escenarios y modos de acceder a la información en la web.

---

<sup>70</sup> El código y los datos liberados por ProPublica para el análisis de Machine Bias puede consultarse en el siguiente repositorio en la plataforma GitHub: <https://github.com/propublica/compas-analysis>

<sup>71</sup> Bradshaw, P. (2013). *Scraping for Journalists: How To Grab Data From Hundreds of Sources, Put It In a Form You Can Interrogate and Still Hit Deadlines*. Canada: Leanpub.

<sup>72</sup> Datajournalism.com. (s.f.). *Data scraping for stories*, Datajournalism.com. Consultado a 12 de septiembre de 2021, disponible en: <https://datajournalism.com/read/newsletters/data-scraping-for-stories>

- Algunas páginas web presentan además sistemas de autenticación que impiden la conexión automática a la web desde una secuencia de comandos. Para ello, el profesional de la información utiliza emuladores como Chrome Headless que simulan el comportamiento humano a la hora de acceder a una web.
- Los cambios recurrentes en las páginas web también provocan que las secuencias de comandos programadas para extraer datos de una web fallen con el tiempo.
- Los bloqueos de IP del usuario, en este caso periodista, que pretende extraer datos son comunes por parte de ciertas webs y organismos. Los bloqueos de IP se producen cuando el periodista accede a una web múltiples veces en un período corto de tiempo desde el mismo ordenador. Estas zancadillas pueden resolverse cambiando la IP de acceso a la web periódicamente o pausando las consultas y accediendo a la información de forma paulatina y en diferentes espacios de tiempo. De esta forma, el acceso a la web no es considerado como automatizado por lo que el robot programado por el periodista puede seguir recopilando información sin ser detectado.
- Por último, otro punto a tener en cuenta es la legalidad del acceso y descarga de información disponible en la web pero que no está diseñada para ser recopilada <sup>73</sup>. El periodista debe cubrirse las espaldas y trabajar con un equipo legal en caso de que los datos extraídos sean utilizados en alguna publicación.

### **3.4.1.2. Bots para la ejecución de tareas**

El objetivo de este tipo de bots es facilitar algunas de las tareas que se realizan en el día a día en una reacción. Para entender mejor sus objetivos y su configuración se exponen a continuación algunos ejemplos relevantes.

---

<sup>73</sup> Ibid.

## ***Klaxon***

Con el uso extendido de aplicaciones de mensajería como Slack en los medios de comunicación, algunas publicaciones han desarrollado *bots* de uso interno para seguir diferentes temas.

*Klaxon*<sup>74</sup> es un *software* de licencia libre y gratuito creado y desarrollado en la redacción del medio estadounidense especializado en la cobertura de prisiones y delitos, *The Marshall Project*. La publicación contó con el apoyo de la organización periodística sin ánimo de lucro *Knight-Mozilla OpenNews* para la creación de esta herramienta.

*Klaxon* es un robot *open source* que, una vez instalado en Slack, alerta a los profesionales de la información de los cambios que se producen en una página web. Para *The Marshall Project* y decenas de medios, *Klaxon* ha supuesto un cambio en el modo de acercarse a las historias que cubren en el día a día. Tras observar y registrar los cambios que se producen en una página web, *Klaxon* ha guiado a los periodistas en la cobertura de diferentes historias o investigaciones. El trabajo automatizado de *Klaxon* ha proporcionado a los periodistas el tiempo suficiente para reaccionar a esos cambios en las páginas web e investigarlos en profundidad<sup>75</sup>.

## ***FOIAMachine***

FOIAMachine es otra herramienta creada con el objetivo de ayudar al periodista a monitorear las solicitudes de acceso a la información que presenta en los organismos públicos para obtener datos o información.

---

<sup>74</sup> The Marshall Project. (s.f.). *Klaxon*. Consultado a 12 de septiembre de 2021, disponible en: <https://newsklaxon.org/>

<sup>75</sup> *Klaxon* es una herramienta *open source*. El proyecto y código de programación que potencia *Klaxon* está disponible en la plataforma GitHub: <https://github.com/themarshallproject/klaxon>

*FOIAMachine*<sup>76</sup> es también una herramienta open source, creada por la organización sin ánimo de lucro MuckRock, que permite a los periodistas enviar, organizar y compartir solicitudes de acceso a la información para la obtención de datos y documentos de los organismos gubernamentales. Este *bot* es especialmente útil para crear solicitudes de información de forma rápida, ya que cuenta con los contactos de más de 500 oficinas del Gobierno estadounidense, y para mantener un registro de las respuestas a las solicitudes.

### **3.4.2. Bots públicos**

#### **3.4.2.1. Bots conversacionales**

En el sector del periodismo, los bots conversacionales o chatbots son aquellas máquinas automatizadas que interactúan con el lector para ofrecer una información. Los bots conversacionales pueden desarrollarse como una pieza informativa en el contexto del medio de comunicación o como un servicio que ofrece el medio en las redes sociales. A continuación se examinan algunos chatbots publicados recientemente para demostrar las características de este tipo de bots.

#### ***Valtteri, the Election Bot***

Uno de los objetivos del proyecto Immersive Automation, un grupo de estudio creado por profesores de la Swedish School of Social Science y del VTT Technical Research Centre of Finland, es el análisis de la automatización en los procesos editoriales<sup>77</sup>.

---

<sup>76</sup> *FOIAMachine* es una herramienta open source. El proyecto y código de programación que potencia *FOIAMachine* está disponible en la plataforma *GitHub*: <https://github.com/cirlabs/foiamachine>

<sup>77</sup> Immersive Automation. (2018). *Immersive Automation: summary of project outcomes*. Immersive automation. Consultado a 15 de septiembre de 2021, disponible en: <http://immersiveautomation.com/wp-content/uploads/2018/10/Immersive-Automation-for-web-page.pdf>

Por esta razón, con el fin de aprender las dificultades alrededor de la automatización e identificar qué procesos merece la pena automatizar, Immersive Automation, creó un robot para informar sobre contenido electoral al que bautizaron como Valtteri <sup>78 79</sup>.

Según Immersive Automation, los elementos y capacidades con las que dotaron a su bot Valtteri son:

- La capacidad de análisis de datos e informaciones y el desarrollo de diferentes criterios para identificar qué es noticia.
- La capacidad de escribir gracias a la creación y uso de diferentes plantilláis automatizadas.

Una vez programadas estas capacidades en Valtteri, el bot fue capaz de combinar los hechos relevantes que observaba al analizar datos con los textos escritos previamente por los reporteros que trabajaron en el proyecto y que atendían los diferentes escenarios que pueden ocurrir durante la campaña electoral. Una vez combinados estos elementos, Valtteri fue capaz de generar noticias cortas sobre la última hora electoral.

Una vez creada la lógica del robot, la traducción en idiomas es relativamente sencilla ya que la estructura del robot ya está programada. Immersive Journalism dotó a Valtteri de la capacidad de hablar y escribir en finlandés, sueco e inglés.

---

<sup>78</sup> Immersive Automation. (7 de abril de 2017). *Valtteri the election bot*. Immersive Automation. Immersive automation. Consultado a 15 de septiembre de 2021, disponible en: <http://immersiveautomation.com/valtteri-election-bot/>

<sup>79</sup> Department of Computer Science at the University of Helsinki. (4 de Julio de 2017). *Valtteri the Election Bot generates news in three languages*. University of Helsinki. Consultado a 15 de septiembre de 2021, disponible en: <https://www.cs.helsinki.fi/en/news/86516>

Otro de los puntos que pone de manifiesto Valtteri es la aplicación de estas técnicas y aprendizajes en otros proyectos. Immersive Automation utilizó el sistema creado para Valtteri para analizar datos y estadísticas de crimen e informar al público de ello.

## **Un bot como asistente de finanzas personal**

Por su parte, *The Washington Post* desarrolló en 2020 un asistente de finanzas personal <sup>80</sup> con una mayor capacidad de interactuar con el lector. Para ello, creó un alter ego en forma de bot de una de sus columnistas especializadas en finanzas personales, Michelle Singletary.

El bot creado por *The Washington Post* fue programado teniendo en cuenta diferentes escenarios y preguntas que los lectores podrían tener sobre finanzas. De nuevo, el diseño previo y la predicción de los escenarios en los que el lector se puede encontrar o la preguntas que puede tener son claves para el correcto funcionamiento del bot una vez publicado.

El bot interactúa con el lector, le hace una serie de preguntas para recabar la información necesaria con el fin de ofrecerle una serie de escenarios posibles y soluciones. Al comienzo de la conversación, el chatbot pide al lector que tenga a mano sus resguardos del banco de la hipoteca, del alquiler del coche y la información del plan de su jubilación.

Además, el bot advierte al lector de que va a hacerle una serie de preguntas para determinar su situación financiera:

“Mientras avanzas respondiendo a las preguntas, puedes encontrarte que estás en una posición mucho más relajada a nivel financiero de lo que pensabas. O puede que tengas que afrontar la verdad dura de tu estado financiero. Lo más importante es que no hago juicios. Los números

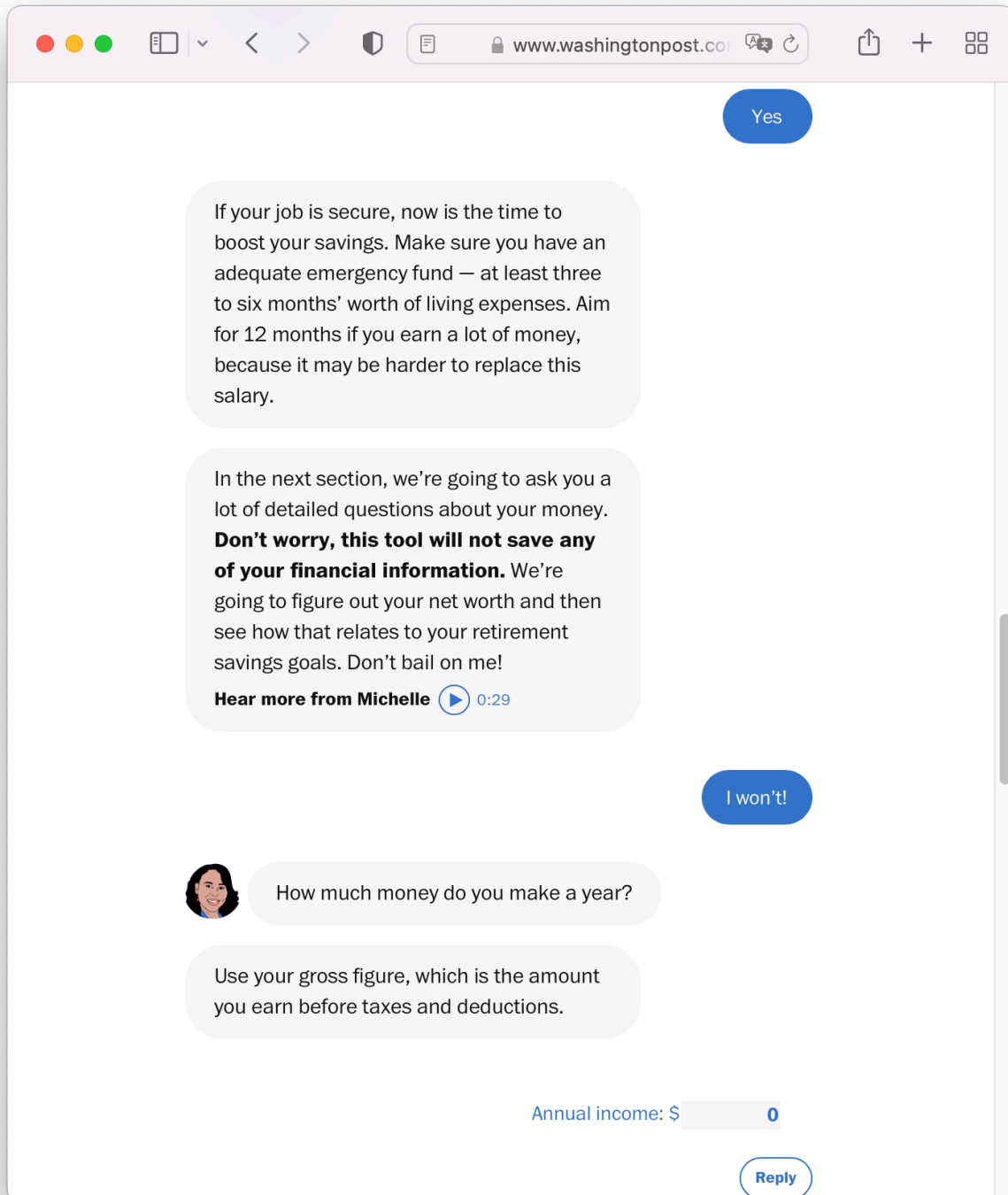
---

<sup>80</sup> Singletary, M. y Shin, Y. (22 de mayo de 2020). *Ask your retirement questions to our Michelle Singletary Bot*. *The Washington Post*. Consultado a 30 de mayo de 2020, disponible en: <https://www.washingtonpost.com/graphics/2020/business/retirement-planning-bot/>

simplemente dicen en qué situación estás a día de hoy. Ya seas joven o mayor, si no te gusta lo que ves, siempre puedes cambiar tu situación.”

El chatbot además incluye notas de voz grabadas por la periodista de *The Washington Post* que ofrecen contexto a las conclusiones que el lector recibe al introducir su información. Este proyecto es una forma de hacer más cercana la información al público, gracias a que el contenido y conversación con el chatbot varía en función de los detalles financieros y circunstancias de cada lector.

Figura 7. Reproducción de una conversación con el lector del bot conversacional y asistente personal de finanzas desarrollado por *The Washington Post*.



Fuente: *The Washington Post*.

### 3.4.2.2. Bots como herramienta informativa

Se conoce como bot informativo a aquel robot que produce informaciones relevantes de manera automática. Este tipo de bots pueden encontrarse como piezas periodistas entre la oferta de los medios de comunicación. A continuación, se expone un caso como ejemplo de este tipo de bots cuyo principal objetivo es informar al lector.

Quakebot <sup>81</sup> es un bot en forma de artículo publicado Los Angeles Times que ofrece información en tiempo real de los terremotos que se producen en California. Quakebot es un robot que ofrece información actualizada del epicentro del terremoto, su escala de Richter y el área afectada. Ésta es una información automatizada de servicio público que ofrece un primer acercamiento para el lector al hecho noticioso y al peligro al que se expone en una zona sísmica.

La cobertura de información de urgencia para los lectores por Quakebot ha supuesto una fuente de ingresos adicional para Los Angeles Times debido al número de personas que se han unido como suscriptores tras leer estas informaciones <sup>82</sup>.

Los Angeles Times destaca en este tipo de contenidos automatizados que ofrecen una información breve pero precisa de un evento informativo que ha ocurrido unos segundos previos a la publicación de la noticia. Es el caso también de California Fire Map <sup>83</sup>, un artículo automatizado en forma de mapa interactivo que informa de los incendios que surgen en tiempo real en California.

---

<sup>81</sup> Los Angeles Times. (s.f.). *Quakebot*. Los Angeles Times. Consultado a 6 de enero de 2022, disponible en: <https://www.latimes.com/people/quakebot>

<sup>82</sup> Greenway, R. (20 de abril de 2020). *Tools for Public Service journalism: Q&A with Casey Miller*. RJI. Consultado a 6 de enero de 2022, disponible en: <https://rjionline.org/news/tools-for-public-service-journalism-qa-with-casey-miller/>

<sup>83</sup> Los Angeles Times. (s.f.). *California fire map: Active wildfires*. Los Angeles Times. Consultado a 2 de febrero de 2022, disponible en: <https://www.latimes.com/wildfires-map/>

California Fire Map lee y analiza ocho fuentes de datos e información diferentes para determinar qué incendios están activos en el estado de California. Una vez identificados los incendios y calculados sus perímetros, California Fire Map visualiza en un mapa interactivo el incendio, el área afectada y ofrece información adicional contrastada del evento noticioso. Además, la información recabada por el bot se reutiliza en otras piezas informativas del medio argelino para ofrecer contexto.

### **3.4.2.3. Bots en redes sociales**

Además de los bots desarrollados en las páginas nativas de los medios, Telegram, Twitter, Facebook Messenger y WhatsApp son algunas de las plataformas sociales que ofrecen un buen ecosistema para el desarrollo de la automatización en forma de robots reporteros.

Pese a que los bots con objetivos periodísticos tuvieron una gran acogida en Telegram y Facebook Messenger hace unos años, Twitter es la plataforma que ha mantenido un mayor número de bots con fines periodísticos a lo largo de tiempo.

Pese a la gran cantidad de bots maliciosos en Twitter <sup>84</sup> con diferentes objetivos como el reciente intento de participar en el discurso digital sobre las vacunas de la COVID-19 <sup>85</sup>, la red social de los 280 caracteres es la plataforma que más flexibilidad ofrece para la creación de bots con objetivos periodísticos. La API de Twitter es una de las más flexibles para la publicación de contenidos lo que permite a los periodistas adaptar sus ideas y sus proyectos de automatización a la plataforma social.

---

<sup>84</sup> Confessore, N., Dance, G., Harris, R. y Hansen, M. (27 de enero de 2018). *The follower factory*. The New York Times. Consultado a 1 de junio de 2019, disponible en: <https://www.nytimes.com/interactive/2018/01/27/technology/social-media-bots.html>

<sup>85</sup> Polaino, R. C., Cárdbaba, M., Cirujano, E. y Cambra, U. (2021). *Las vacunas contra la Covid en Twitter: redes, mensajes y bots*. In *Digital media: el papel de las redes sociales en el ecosistema educomunicativo en tiempos de Covid-19*. McGraw-Hill Interamericana de España, 487-498.

De esta flexibilidad y facilidad para la creación de bots periodísticos en Twitter se han beneficiado los medios de comunicación, creando proyectos con el objetivo de llegar e interactuar con nuevas audiencias en la red social.

*The Next To Die* <sup>86</sup> es un bot desarrollado por *The Marshall Project* que informa de las fechas de la condena de los presos sentenciados a pena de muerte en los diferentes estados de Estados Unidos. El proyecto está publicado tanto como pieza periodística dentro de la oferta informativa del medio como en la red social Twitter <sup>87</sup> para facilitar que los usuarios de la red sigan las actualizaciones del proyecto.

Por otro lado, *censusAmericans* <sup>88</sup> es un bot desarrollado por el medio estadounidense *FiveThirtyEight* en la plataforma social Twitter. El bot publica un tuit cada hora con información anónima sobre personas individuales que viven en Estados Unidos. Los datos de los que se alimenta *censusAmericans* provienen de la oficina del Censo de Estados Unidos. La metodología usada para construir el bot describe que una serie de secuencias de código escrito en Python reconstruyen una breve biografía anónima por cada estadounidense utilizando los datos del censo.

---

<sup>86</sup> The Marshall Project. (s.f.). *The next to die*. The Marshall Project. Consultado a 11 de enero de 2022, disponible en: <https://www.themarshallproject.org/next-to-die>

<sup>87</sup> La versión en Twitter del bot *The Next To Die* puede consultarse en el siguiente enlace: <https://twitter.com/thenexttodie>

<sup>88</sup> Zhang, J. (24 de Julio de 2015). *Introducing Censusamericans, a Twitter bot for America*. FiveThirtyEight. Consultado a 11 de enero de 2022, disponible en: <https://fivethirtyeight.com/features/introducing-censusamericans-a-twitter-bot-for-america/>

### 3.5. Inteligencia artificial y *Newsmaking*

La automatización de procesos ha llegado al periodismo y, en concreto, a la generación y creación de noticias, conocida en inglés como *newsmaking* <sup>89</sup>. Cada vez son más los medios que han desarrollado proyectos para automatizar la redacción de sus noticias e informaciones.

García-Alonso <sup>90</sup> hace referencia al periodista e investigador, Tom Van de Weghe, que durante su estancia en 2019 en la Universidad de Stanford con una beca *John S. Knight Journalism Fellowship* se entrevistó con líderes de innovación en medios como *The Associated Press* o *The Washington Post*. Van de Weghe afirma que durante esas entrevistas pudo conocer de primera mano como un buen número de medios relevantes ya han integrado la inteligencia artificial en sus operaciones de noticias <sup>91</sup>.

La iniciativa de generar noticias por ordenador, que comenzó como un experimento, ya es una realidad en buena parte de las redacciones más importantes del mundo como se ha visto en la Introducción de esta tesis con la iniciativa *Project Cyborg* de *Bloomberg* y el robot *Heliograf* de *The Washington Post*. Como se ha expuesto previamente, las secciones o temas en los que se han hecho más avances en pos de la automatización de la redacción de contenidos son la economía, las finanzas y los eventos deportivos.

¿Por qué son estos los principales temas donde se ha visto un mayor desarrollo de este tipo de técnicas? La principal razón es que estas coberturas periodísticas suelen seguir

---

<sup>89</sup> Marconi, F. (2020). *Newsmakers: Artificial Intelligence and the Future of Journalism*. United States: Columbia University Press.

<sup>90</sup> García-Alonso, P. (2020). “Claves del Periodismo Emergente en el diario *The Washington Post*”, pp. 85-87, con Jaspe Nieto, Javier, Capítulo en: *Periodismos Emergentes: Transformación y Revitalización del Periodismo en la Era Digital*. 2020, Universidad Complutense de Madrid, 179 págs.

<sup>91</sup> Laboratorio de Periodismo (2020). “Guía para introducir la inteligencia artificial en las redacciones”, 08.06.2020 Fundación Luca de Tena, en <https://laboratoriodeperiodismo.org/guia-para-introducir-la-inteligencia-artificial-en-las-redacciones/>

patrones muy similares. Por la naturaleza de las mismas, los profesionales de la información pueden predecir la estructura final del texto por lo que pueden programar con relativa facilidad un algoritmo que generará una noticia redactada correctamente y con amplitud de vocabulario. Además, los profesionales de la información pueden dotar al algoritmo de inteligencia artificial para mejorar con el fin de mejorar el producto informativo.

En el apartado de la inteligencia artificial, los avances en el sector del periodismo han permitido la incorporación de nuevas técnicas. El machine learning, el aprendizaje supervisado y el aprendizaje no supervisado, todas ellas técnicas inteligentes para el análisis e interpretación de los datos, también han sido incorporados en los últimos años a diferentes procesos, tareas y objetivos en las redacciones de los medios de comunicación <sup>92</sup>. Los modelos de inteligencia artificial y machine learning permiten a los periodistas analizar datos e identificar patrones y tendencias, además de generar conocimiento.

De la misma forma que los medios de comunicación trabajan en generar noticias escritas por un robot, también existe una corriente que se centra en identificar temas relevantes para la audiencia a los que no se les está dando mucha importancia y que pueden formar parte de coberturas futuras.

Por ejemplo, mediante el uso de una serie de modelos de inteligencia artificial, el diario estadounidense *The Wall Street Journal*, analizó y extrajo conclusiones de los artículos publicados durante la última década. Para ello, utilizaron un método de análisis del lenguaje o procesamiento de lenguaje natural (NLP) denominado *Doc2Vec*.

Este método permitió al equipo de investigación y desarrollo de *The Wall Street Journal (WSJ R&D)* descubrir los vacíos informativos y los temas de relevancia que los editores del diario estaban dejando fuera de su cobertura de la actualidad. Además, la

---

<sup>92</sup> Marconi, F. y Siegman, A. (2017). *The Future of Augmented Journalism: A guide for newsrooms in the age of smart machines*. Associated Press. Disponible en: [https://www.ap.org/assets/files/2017\\_ai\\_guide.pdf](https://www.ap.org/assets/files/2017_ai_guide.pdf)

aplicación de esta técnica de inteligencia artificial, les ayudó a definir cuáles son los temas que más interesan a sus lectores <sup>93</sup>.

El conocimiento sobre la inteligencia artificial se ha democratizado en los últimos años desde que múltiples modelos de machine learning están disponibles en la red para su uso libre y gratuito. Modelos que pueden ser desarrollados fácilmente en las redacciones que cada vez cuentan con más ingenieros y profesionales con conocimientos tecnológicos.

Estas herramientas ayudan a mejorar y perfeccionar los métodos periodísticos, pero nunca los reemplazarán. Según Marconi, la inteligencia artificial puede ayudar en la identificación de historias, pero los periodistas siempre deben aplicar su juicio y opinión editorial <sup>94</sup>.

De acuerdo con HANSEN, M., ROCA-SALES, M., KEEGAN, J.M. y KING <sup>95</sup> las herramientas de inteligencia artificial pueden ayudar a contar nuevas historias que hace una década no se podían contar por razones técnicas.

A pesar de la evolución de la inteligencia artificial en el sector del periodismo en los últimos años, los autores advierten de que todavía existe camino por recorrer ya que la aplicación y los conocimientos de la inteligencia artificial en el sector de la computación son mucho más avanzados que en el sector del periodismo. Este hecho, advierten los autores,

---

<sup>93</sup> Marconi, F. (13 de Octubre de 2019). *How the Wall Street Journal is using Deep Learning to inform content strategy*. Medium. Consultado a 13 de enero de 2022, disponible en: <https://medium.com/the-wall-street-journal/how-the-wall-street-journal-is-using-deep-learning-to-inform-content-strategy-4b4a07090110>

<sup>94</sup> Ibid.

<sup>95</sup> Hansen, M., Roca-Sales, M., Keegan, J.M. y King, G. (2017). *Artificial Intelligence: Practice and Implications for Journalism*. Tow Center for Digital Journalism y Brown Institute for Media Innovation. Disponible en: <https://academiccommons.columbia.edu/doi/10.7916/D8X92PRD>

puede conducir a malas prácticas en el periodismo si la inteligencia artificial se aplica sin los conocimientos adecuados.

Por último, debido al carácter impredecible de la inteligencia artificial, los autores no se atreven a predecir qué problemas surgirán en el futuro. Por ello, proponen que la aplicación de estas nuevas técnicas sea revisadas y vigilada editorialmente por los periodistas y los profesionales de la información <sup>96</sup>.

A pesar de la incertidumbre y la dudas en el uso de estas técnicas de inteligencia artificial, sus aplicaciones en el sector del periodismo dibujan un crecimiento exponencial en los próximos años.

De hecho, Marconi y Siegman predicen que en 2027, los medios tendrán un arsenal de herramientas propulsadas por la inteligencia artificial, y los periodistas utilizarán sin apenas percatarse robots inteligentes en sus día a día.

De acuerdo con los autores, el machine learning será capaz de producir noticias más elaboradas de forma automática en el próximo lustro. Además, según Marconi y Siegman, la inteligencia artificial permitirá a los reporteros analizar datos, identificar tendencias combinando múltiples fuentes de datos, descubrir conclusiones que no podrían descubrir por sí mismos, convertir datos y conversaciones en texto, transformar textos en audios y vídeos, entender sentimientos, analizar imágenes reconociendo objetos, textos y colores y mucho más <sup>97</sup>.

---

<sup>96</sup> Ibid.

<sup>97</sup> Marconi, F. y Siegman, A. (2017). *A day in the life of a journalist in 2027: Reporting meets AI*. Columbia Journalism Review.

Disponible en: <https://www.cjr.org/innovations/artificial-intelligence-journalism.php#:~:text=By%202027%2C%20newsrooms%20will%20have,out%20straightforward%2C%20automated%20news%20reports>.



## **CAPÍTULO IV: Automatización en el periodismo de investigación**



**L**a relación entre la automatización, los algoritmos y el periodismo de investigación es estrecha en la concepción actual de la disciplina. Ya sea para analizar datos mediante el uso de algoritmos o para investigar las implicaciones del uso de esos algoritmos en la sociedad, el periodista ha asimilado las nuevas tecnologías para la ejecución de estas tareas.

Un algoritmo, o sistema de decisión automática, es una secuencia o serie de reglas, operaciones y cálculos ejecutadas por un ordenador para resolver un problema. Hoy en día, aunque no siempre se perciban a simple vista, la vida de las personas se rige en un aspecto u otro por algoritmos.

En su versión más simplificada, un algoritmo es una lista de instrucciones. Un algoritmo funciona como una receta. Por ejemplo, un algoritmo muy sencillo que no implica ningún elemento de inteligencia artificial consistiría en el análisis automático de un conjunto de datos. Al igual que ocurre al seguir una receta, la calidad del plato final puede verse afectada por diferentes razones.

Las instrucciones de la receta pueden ser incorrectas o poco específicas, de la misma forma que un algoritmo puede estar mal diseñado. Los ingredientes de la receta pueden estar

mal escogidos, al igual que el algoritmo puede depender de otras librerías de código que no son las mejores para realizar la tarea encomendada. Una receta puede estar cocinada con ingredientes en mal estado, de la misma manera que un algoritmo puede servirse de datos erróneos o inadecuados.

Todos y cada uno de estos elementos contribuyen al resultado final de la receta o, en este caso, a las conclusiones del algoritmo. Por ello, una parte fundamental de la investigación periodística sobre algoritmos se fundamenta en identificar y entender los diferentes pasos implicados en la ejecución del algoritmo.

A la hora de trabajar con algoritmos desde un punto de vista periodístico, se deben considerar dos ramas diferentes. Por un lado, el profesional de la información puede analizar, e investigar los algoritmos desarrollados por terceros como administraciones o empresas privadas. Por otro lado, el profesional de la información puede desarrollar sus propios algoritmos para que le ayuden a realizar sus labores diarias, analizar datos o investigar otros algoritmos o procesos.

Sin embargo, como indica O'NEIL, “no es un trabajo sencillo”<sup>98</sup> reducir el conocimiento, el desempeño, la interpretación y el potencial de los humanos a un conjunto de reglas como las de un algoritmo.

Una de las advertencias que hace O'NEIL es que los algoritmos se alimentan en la mayoría de casos de datos. Por ende, por muy bien diseñado que esté un algoritmo, si los datos no son correctos o contienen errores, el resultado y la decisión que tome el algoritmo será errónea. La autora cita además una frase muy utilizada en el sector de la ciencias de datos para describir el procesamiento y las conclusiones a las que llegan los algoritmos cuando los datos son incorrectos como es “garbage in, garbage out” (“basura dentro, basura fuera”).<sup>99</sup>

---

<sup>98</sup> O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. United Kingdom: Crown, p. 12.

<sup>99</sup> Ibid, p. 126.

Con el fin de conocer mejor qué es un algoritmo y qué usos tiene en el sector del periodismo de investigación, esta tesis analiza las dos aplicaciones fundamentales de los algoritmos en una redacción.

En primer lugar, esta disertación estudia cómo los periodistas, en particular aquellos que trabajan en investigaciones o cubren tecnología, analizan los algoritmos presentes en la sociedad y su repercusión en la vida de las personas.

En segundo lugar, se expone cómo los propios profesionales de la información pueden beneficiarse del desarrollo y el uso de algoritmos para mejorar su productividad, analizar datos o contar historias que de otra forma no se podrían contar.

#### **4.1. Análisis e investigación de algoritmos**

Las piezas periodísticas que detallan el uso de los algoritmos por parte de gobiernos, administraciones o empresas privadas se han hecho un hueco en las portadas digitales y las páginas de papel de los principales diarios internacionales. A la vez que los algoritmos se han convertido en un elemento habitual en la vida de las personas, el interés en su uso ha crecido en la opinión pública y en el periodismo de investigación.

Como se ha definido anteriormente, un algoritmo es una serie de reglas, operaciones y cálculos con un objetivo concreto. Cabe destacar que el diseño y programación de esas reglas puede afectar al resultado final que el algoritmo produce. De esta forma, existen una serie de implicaciones y connotaciones que siempre se deben tener en cuenta a la hora de analizar un algoritmo.

Aunque la disciplina periodística fue una de las primeras en preguntarse sobre la ética y el funcionamiento de estos algoritmos, lo cierto es que el interés ha trascendido a otras disciplinas y ámbitos. Los artículos e investigaciones publicados sobre algoritmos y la presión hecha por expertos y profesionales de la ciencia de datos ha hecho que algunas administraciones públicas reaccionen. En 2018, el entonces alcalde de Nueva York, Bill de

Blasio, anunció un grupo de trabajo para evaluar las aplicaciones de los algoritmos en uso por los organismos de la ciudad <sup>100</sup>.

Al otro lado del Atlántico, y de acuerdo con el informe Automating Society Report 2020, publicado por el centro de estudios europeo sin ánimo de lucro, Algorithm Watch, existen una serie de riesgos al analizar los procesos de toma de decisiones automatizados (ADM) o algoritmos que a día de hoy están en funcionamiento en Europa <sup>101</sup>. El estudio destaca la opacidad de la mayoría de esos sistemas o algoritmos frente al beneficio o la utilidad social de los mismos <sup>102</sup>. El informe detalla que “cuando se mira el estado actual de los sistemas ADM en Europa, los ejemplos positivos con claros beneficios son poco comunes y la gran mayoría de los usos tienden a poner en riesgo a las personas en lugar de ayudarlas”.

El debate y el análisis de los algoritmos también ha dado el salto a la literatura académica en los últimos años. Algorithmic journalism o periodismo de algoritmos es un término que pretende describir los procedimientos que los recientes cambios tecnológicos, en forma de algoritmos, han traído al campo del periodismo.

Al igual que ocurre con el resto de especializaciones, descritas en capítulos anteriores, no existe un consenso en la conceptualización del periodismo de algoritmos. De hecho, diferentes autores ofrecen conceptos diferentes, más y menos amplios, en su cobertura de campo del periodismo de algoritmos.

---

<sup>100</sup> Nyc.gov. (16 de mayo de 2018). *Mayor de Blasio Announces First-In-Nation Task Force To Examine Automated Decision Systems Used By The City*, *The Official Website of the City of New York*, NYC. Disponible en: <https://www1.nyc.gov/office-of-the-mayor/news/251-18/mayor-de-blasio-first-in-nation-task-force-examine-automated-decision-systems-used-by>

<sup>101</sup> Algorithm Watch. (2020). *Automating Society Report 2020*, Algorithm Watch. Disponible en: <https://automatingsociety.algorithmwatch.org/wp-content/uploads/2020/12/Automating-Society-Report-2020.pdf>

<sup>102</sup> OdiseIA. (18 de noviembre de 2020). *Automating society report 2020. Algoritmos en el poder; opacidad frente a utilidad social*. Medium. Consultado a 24 de enero de 2021, disponible en: <https://medium.com/odiseia/automating-society-report-2020-algoritmos-en-el-poder-opacidad-frente-a-utilidad-social-d9e78319cb2d>

Una de las primeras concepciones del periodismo de algoritmos abarca diferentes elementos y tareas. Para Hamilton y Turner, el periodismo de algoritmos supone el uso de algoritmos, datos y conocimientos de las ciencias sociales con el objetivo de ejercer la función de rendición de cuentas propia del periodismo <sup>103</sup>. Según Kotenidis, ésta es una definición general e inclusiva de la amplia gama de aplicaciones que tiene el periodismo de algoritmos <sup>104</sup>.

Una definición posterior, como la de Graefe, equipara los términos de periodismo de algoritmos y de periodismo automatizado. Su definición conceptual es más contenida. El autor fundamenta que el periodismo de algoritmos es la disciplina que permite producir informaciones basadas en datos limpios, estructurados y fiables <sup>105</sup>. Incluso, puntualiza que estos contenidos informativos se pueden personalizar a las necesidades del lector de forma rápida, barata y eficaz — cometiendo menos errores de los que podría cometer un humano <sup>106</sup>.

Kotenidis y Veglis exponen además la corriente de que el periodismo de algoritmos comparte características y técnicas con el periodismo de datos o periodismo asistido por ordenador. En cambio, puntualizan que, mientras el periodismo de datos se centra en el uso mundano de la tecnología en labores como almacenar, acceder o analizar datos, el periodismo de algoritmos busca analizar y desarrollar las capacidades de procesamiento del software más moderno <sup>107</sup>.

---

<sup>103</sup> Hamilton, J. T. y Turner, F. (2009). *Accountability through Algorithm: Developing the Field of Computational Journalism*. Center for Advanced Study in the Behavioral Sciences Summer Workshop. Disponible en: <http://web.stanford.edu/~fturner/Hamilton%20Turner%20Acc%20by%20Alg%20Final.pdf>

<sup>104</sup> Kotenidis, E. y Veglis, A. (2021). *Algorithmic Journalism—Current Applications and Future Perspectives*, *Journalism and Media.*, 2021; 2(2):244-257. Disponible en: <https://doi.org/10.3390/journalmedia2020014>

<sup>105</sup> Graefe, A. (2016,). Guide to automated journalism. *Columbia Journalism Review*. Consultado a 4 de septiembre de 2021, disponible en: [https://www.cjr.org/tow\\_center\\_reports/guide\\_to\\_automated\\_journalism.php](https://www.cjr.org/tow_center_reports/guide_to_automated_journalism.php)

<sup>106</sup> Ibid

<sup>107</sup> Ibid.

Diakopoulos además ha acuñado y definido el término *algorithmic accountability reporting* <sup>108</sup> que puede traducirse como la cobertura informativa de los algoritmos con el objetivo de rendir cuentas sobre sus efectos en la sociedad.

La cobertura informativa de los algoritmos busca identificar los algoritmos en los que la sociedad no repara pero que rigen o influyen en el día a día de sus vidas. El periodista, por tanto, investiga y estudia las características e impacto de estos algoritmos en la sociedad. El objetivo de estas coberturas o investigaciones es la rendición de cuentas de los actores públicos o privados que han diseñado el algoritmo o han promovido su implantación.

Uno de los principales problemas a los que el periodista se enfrenta al investigar el uso e impacto de los algoritmos utilizados e implantados por terceros es la falta de transparencia en el diseño de las acciones del algoritmo y en el código de programación que permite su funcionamiento.

Ante la falta de transparencia, de acuerdo con Diakopoulos, el elemento clave para realizar con maestría esta especialización dentro del periodismo de investigación es la ingeniería inversa del algoritmo <sup>109</sup>. Es decir, mediante el uso de técnicas computacionales e indagaciones adicionales propias del periodismo de investigación, el periodista tratará de describir los procesos que realiza un algoritmo comenzando por el resultado final del mismo. Su objetivo será desvelar los fundamentos de diseño del algoritmo para entender su funcionamiento <sup>110</sup>.

El diseño de un algoritmo es una de las tareas fundamentales a la hora de crear un algoritmo. Cómo se diseña el algoritmo y qué datos, pasos y casuísticas se tienen en cuenta

---

<sup>108</sup> Diakopoulos, N. (2017). *Algorithmic Accountability Reporting: On the Investigation of Black Boxes*, Tow Center for Digital Journalism, Columbia University. Disponible en: <https://doi.org/10.7916/D8ZK5TW2>

<sup>109</sup> Ibid.

<sup>110</sup> Ibid

para ello, condicionarán el resultado y conclusiones que el algoritmo obtenga una vez ejecutado. De esta forma, conocer quién ha diseñado el algoritmo y si lo ha hecho en base a criterios globales e inclusivos de toda la población es clave para determinar el impacto que las conclusiones del algoritmo puedan tener en la sociedad.

Por otro lado, existen algoritmos que toman decisiones en función de los datos que analizan. El contenido de esas bases de datos influirá en el análisis y resultado final del algoritmo. Por ello, es importante conocer qué datos alimentan el análisis del algoritmo para determinar si el proceso es correcto, ético e inclusivo.

Uno de los primeros medios que comenzó a interesarse por las repercusiones y los valores éticos del uso de algoritmos fue el medio estadounidense especializado en periodismo de investigación, ProPublica. Los reporteros de ProPublica han publicado extensas investigaciones sobre el uso de los algoritmos por parte de los gobiernos y las empresas privadas y el impacto de los mismos en la sociedad. ProPublica sentó las bases prácticas de la investigación de los algoritmos en su serie de reportajes de investigación titulada Machine Bias.

En el reportaje Machine bias in criminal sentencing, ProPublica profundiza en el uso de algoritmos en los juzgados de Estados Unidos para determinar las probabilidades de que una persona detenida por un delito menor reincida y cometa un nuevo acto criminal. La probabilidad calculada por el algoritmo se usa para determinar la sentencia del preso. ProPublica descubrió que el algoritmo estaba sesgado en contra la población negra y calculaba mayores riesgos de reincidencia <sup>111</sup>.

En la investigación Amazon Says It Puts Customers First. But Its Pricing Algorithm Doesn't, los reporteros de ProPublica investigaron el algoritmo de búsqueda y fijación de

---

<sup>111</sup> Larson, J. y Angwin, J. (23 de mayo de 2016). *Machine bias*. ProPublica. Consultado a 13 de febrero de 2019, disponible en: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

precios de Amazon <sup>112</sup>. Tras analizar 250 de los artículos más vendidos, repararon en que Amazon no siempre posiciona el artículo más barato en la primera posición de búsqueda. En cambio, los reporteros de ProPublica descubrieron que el algoritmo de búsqueda de Amazon posicionaba mejor artículos más caros.

#### **4.1.1. Un caso de éxito: *The Markup* y la investigación de algoritmos**

Tras unos inicios tormentosos por discrepancias entre sus cofundadores <sup>113</sup> <sup>114</sup>, *The Markup* echa andar a principios de 2020. *The Markup* es un medio estadounidense fundado, entre otros, por la ex periodista de ProPublica, Julia Angwin.

A comienzos de 2022, *The Markup* lo conformaban un equipo de 30 personas de los cuales la mitad son periodistas de datos, reporteros e investigadores centrados en desvelar los secretos de la tecnología. De hecho, un porcentaje elevado de la redacción que forma *The Markup* a día de hoy trabajó previamente para *ProPublica*.

La principal diferencia de *The Markup* con *ProPublica* es que sus contenidos se centran única y exclusivamente en la tecnología y en el uso que tanto instituciones públicas como privadas hacen de ella. La relación creciente entre el periodismo y la tecnología ha posibilitado que surjan este tipo de cabeceras monotemáticas.

---

<sup>112</sup> Mattu, S. y Angwin, J. (20 de septiembre de 2016). *Amazon says it puts customers first, but its pricing algorithm doesn't*. ProPublica. Consultado a 13 de febrero de 2019, disponible en: <https://www.propublica.org/article/amazon-says-it-puts-customers-first-but-its-pricing-algorithm-doesnt>

<sup>113</sup> The Markup es un medio fundado a finales de 2018 por los periodistas Julia Angwin y Jeff Larson. Por discrepancias entre los fundamentos el lanzamiento oficial del medio se retrasa a 2020. The Markup cuenta con capital de diferentes donantes privados. Entre ellos, destaca la donación de 20 millones de dólares de Craig Newmark, fundador del portal web Craigslist y de Craig Newmark Philanthropies. Por aquel entonces también recibió cuantiosas donaciones de John S. and James L. Knight Foundation (2 millones); and additional support from the Ford Foundation and the John D. and Catherine T. MacArthur Foundation.

<sup>114</sup> Tracy, M. (6 de agosto de 2019). *The Markup, a tech news site, reinstalls its fired editor as part of a fresh start*. The New York Times. Consultado a 20 de marzo de 2021, disponible en: <https://www.nytimes.com/2019/08/06/business/media/the-markup-julia-angwin.html>

*The Markup* es un medio de comunicación online estadounidense que “investiga cómo las compañías más poderosas utilizan la tecnología para transformar la sociedad”. En un manifiesto publicado en la web, *The Markup* explica las bases de su periodismo y define a sus reporteros como periodistas cuantitativos:

“*The Markup* es un medio sin ánimo de lucro que investiga cómo compañías poderosas utilizan la tecnología para cambiar nuestra sociedad. Somos un nuevo tipo de medio formado por incomparables periodistas cuantitativos que persiguen que sus las investigaciones basadas en datos que desarrollan tengan un impacto”.

En el mismo documento, también hacen referencia al método y metodologías que siguen para trabajar sus investigaciones y reportajes:

“Nuestro método es científico: creamos bases de datos desde cero, chequeamos nuestras informaciones y mostramos la metodología de nuestros trabajos. A esto lo llamas el método *The Markup*”.<sup>115</sup>

*The Markup* señala que siguen una metodología científica en todos los procesos periodísticos que ejecutan sus periodistas: desde crear bases de datos, chequear y verificar sus fuentes o mostrar su trabajo, procesos y conclusiones. De hecho, en la carta de presentación de su página web puede leerse que denominan a este modelo o modus operandi, el método *The Markup*.

Su directora, Julia Angwin, expone con detalle el método *The Markup*<sup>116</sup> en una carta donde incide y amplía los detalles de la metodología que siguen, basada en el método científico, y que consta de tres partes diferenciadas:

---

<sup>115</sup> The Markup. (s.f.). *About Us – The markup*. The Markup. Consultado a 20 de marzo de 2021, disponible en: <https://themarkup.org/about>

<sup>116</sup> Angwin, J. (25 de febrero de 2020). *A letter from the editor – The Markup*. The Markup. Consultado a 20 de marzo de 2021, disponible en: <https://themarkup.org/2020/02/25/editor-letter-julia-angwin>

- **Construcción de la historia:** los periodistas de *The Markup* realizan las preguntas oportunas y recopilan o construyen las bases de datos que necesitan para comprobar sus hipótesis.
- **Chequeo de las conclusiones a prueba de bala:** los reporteros de *The Markup* verifican las conclusiones de sus historias mediante un riguroso proceso de revisión, ayudándose de e invitando a expertos externos para desafiar sus hallazgos.
- **Publicación del trabajo:** *The Markup* comparte y publica las bases de datos detrás de las historias que investiga, el código que ejecuta el análisis de datos, así como los detalles de los métodos estadísticos utilizados en el análisis. Además, tratan de explicar su metodología en guías metodológicas detalladas.

Otra de las características diferenciales de *The Markup* es la calidad sobre la cantidad. En este sentido, la misión de la publicación es diferente a la de otros medios de comunicación. De acuerdo con sus principios, *The Markup* no tiene como objetivo primordial marcar el ritmo de la actualidad. De esta manera, advierten al lector en su manifiesto de que no espere leer varias noticias breves y llamativas sobre un mismo tema cuyo fin es el de atraer lectores.

Por otro lado, su manifiesto señala que no buscan el clickbait en los titulares y en las conclusiones de sus historias. Por el contrario, tratan de no exagerar o explotar cualquier conclusión curiosa ya que entienden que “toda base de datos tiene sus limitaciones” <sup>117</sup>.

El objetivo que buscan siguiendo estos principios es contribuir a generar confianza en el tipo de periodismo que realizan, cuando, bajo su punto de vista, la confianza en el mundo del periodista está bajo mínimos <sup>118</sup>.

---

<sup>117</sup> Ibid.

<sup>118</sup> Ibid.

Una de las características que hacen diferente a *The Markup*, además del nicho tecnológico que investiga, es que sus publicaciones no son sólo artículos, también ha introducido productos que los lectores pueden utilizar de forma gratuita. Es el caso de *Blacklight*, un inspector de la privacidad en la red que avisa en tiempo real al usuario de si sus datos personales se están viendo comprometidos cuando navega en la red.

De acuerdo con la propia descripción de *The Markup*, *Blacklight* supone un examen forense completo que ilumina la infraestructura del mercado de la vigilancia digital y los algoritmos que se usan para rastrear al usuario.

### **¿Cómo investiga *The Markup* los algoritmos de terceros?**

*Blacklight* es una herramienta muy sencilla desde el punto de vista del usuario. El lector tan sólo tiene que introducir una url o dirección web en el buscador facilitado por *Blacklight*. Una vez introducido y clicado el botón de búsqueda, el conjunto de algoritmos que alimentan la herramienta programada por el periodista de investigación e ingeniero informático, Surya Mattu, se encarga del resto.

De forma automática, los algoritmos de *Blacklight* siguen los pasos pautados por Mattu y ejecutan diferentes tests sobre la web introducida para determinar hasta qué punto la dicha web espía al lector. En primer lugar, el algoritmo comprueba si la web contiene ad trackers o rastreadores de anuncios y, en caso de estar presentes, si estos envían algún tipo de información a compañías publicitarias.

En segundo lugar, la página web verifica si la web introducida por el usuario utiliza cookies para rastrear los movimientos del usuario en la web.

A continuación, *Blacklight* comprueba si la página web identifica de forma invisible desde qué tipo de navegador ha visitado la página el usuario con el fin de determinar si el portal web está tratando de regatear a los bloqueadores de cookies.

Además, Blacklight también es capaz de determinar si la web contiene algún script o alguna secuencia de código que grabe y registre como el usuario interactúa con la página.

Por otro lado, el algoritmo detrás de la herramienta Blacklight explora una acción un tanto controvertida como es la captura y el registro del texto tecleado y escrito por los usuarios en la página sin que necesariamente hayan enviado o autorizado el envío de la información al portal web.

Por último, Blacklight analiza los nexos de unión del portal web con dos de las grandes plataformas digitales como son Facebook y Google. El algoritmo inspecciona si la web comparte datos de los usuarios con Facebook, así como si distribuye datos de los usuarios a Google para que el gigante tecnológico pueda mostrarles publicidad más personal de acuerdo con su historial de navegación.

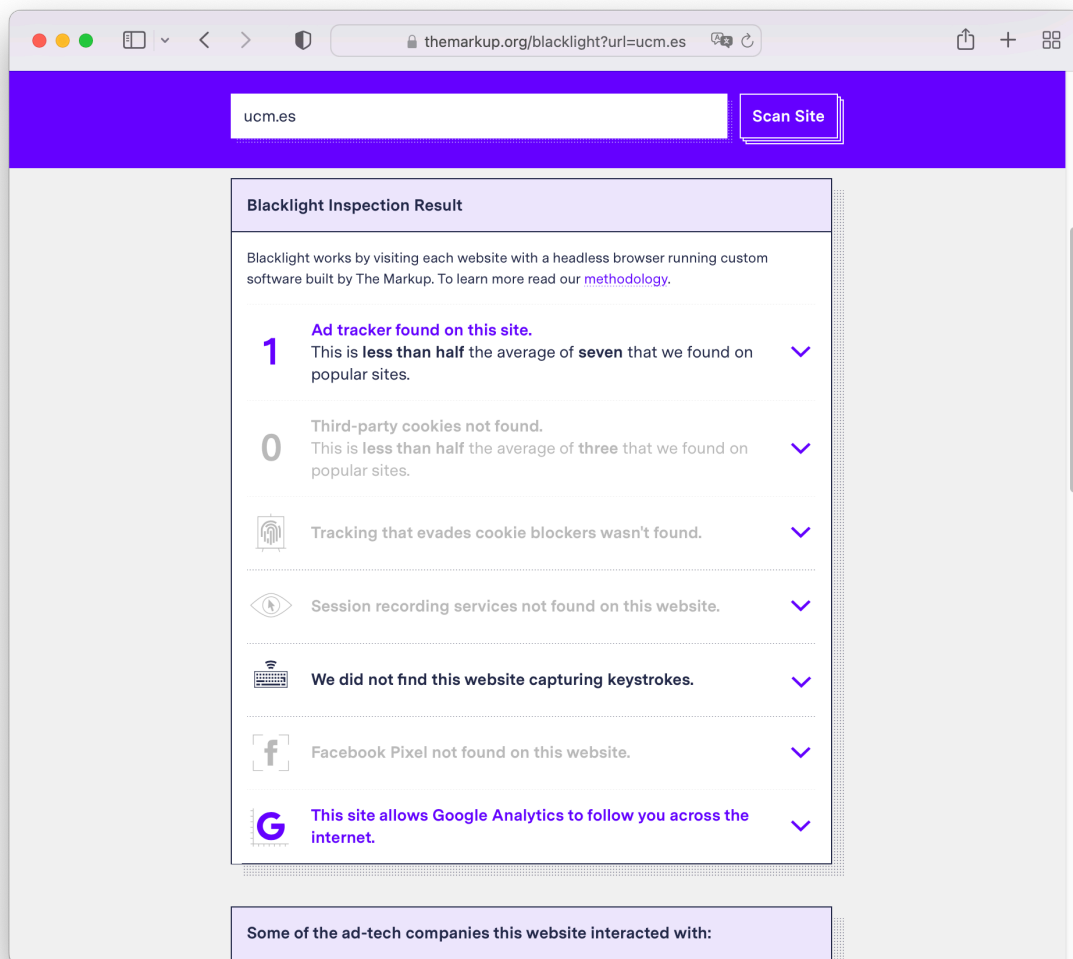
Una vez ejecutados los diferentes exámenes descritos anteriormente que determinan si la web introducida rastrea al usuario, *The Markup* elabora un informe detallado sobre la privacidad del usuario al visitar este portal web y hasta qué punto la web introducida pone en peligro su anonimato.

A continuación, se presentan dos ejemplos de páginas web chequeadas con Blacklight. En primer lugar, se presentan los resultados de la web de la Universidad Complutense de Madrid que apenas espía al lector. En segundo lugar, se exponen los resultados del portal de datos públicos de la COVID-19 del estado de Nevada en Estados Unidos. A pesar de ser un servicio público, esta plataforma es una de las webs más invasivas y que más dato recoge del usuario como relatan los periodistas de *The Markup* en una de sus investigaciones <sup>119</sup>.

---

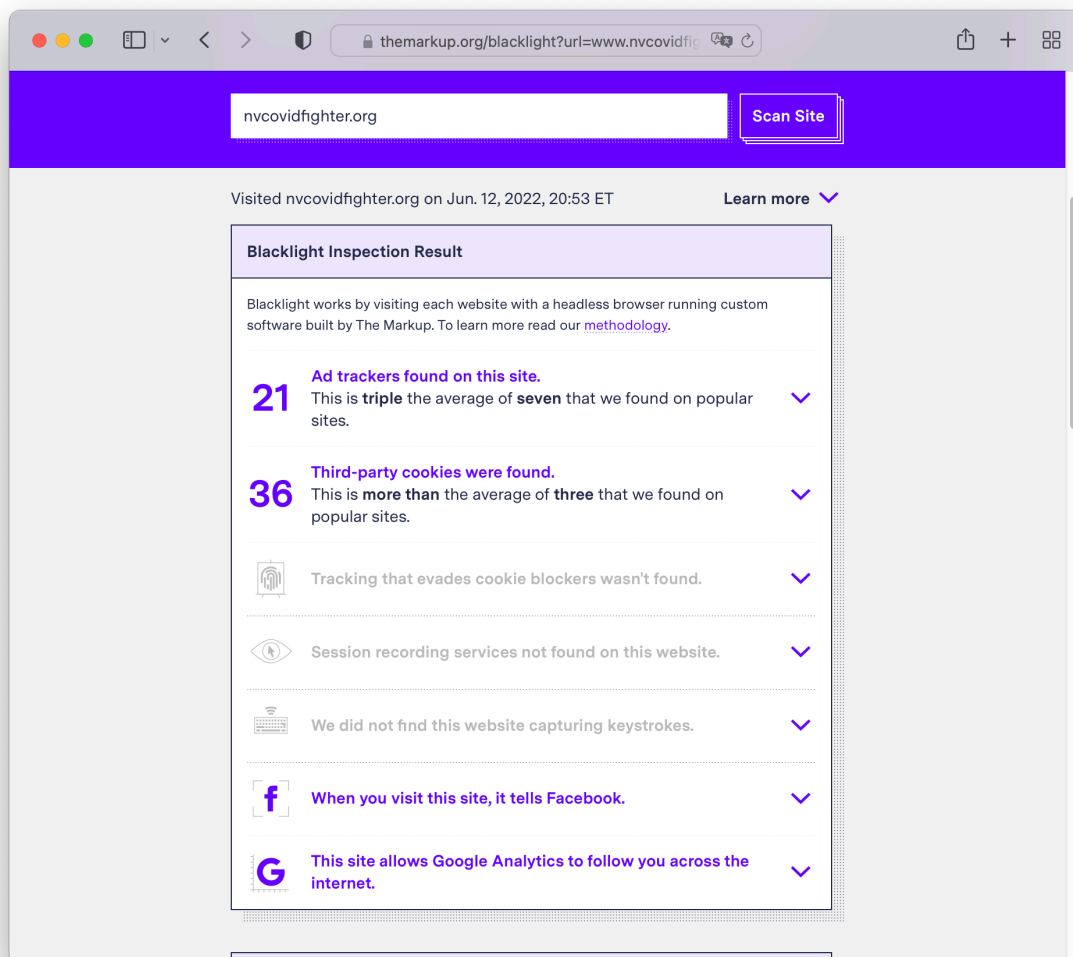
<sup>119</sup> Keegan, J. y Lecher, C. (24 de marzo de 2021). *We ran tests on every state's COVID-19 vaccine website – the markup*. We Ran Tests on Every State's COVID-19 Vaccine Website – The Markup. Consultado a 12 de enero de 2022, disponible en: <https://themarkup.org/blacklight/2021/03/24/we-ran-tests-on-every-states-covid-19-vaccine-website>

Figura 8. Resultado de la búsqueda de la web de la Universidad Complutense de Madrid (ucm.es) en la herramienta *Blacklight*.



Fuente: *The Markup*.

Figura 9. Resultado de la búsqueda del portal de datos de Nevada sobre la COVID-19 en la herramienta Blacklight <sup>120</sup>.



Fuente: *The Markup*.

<sup>120</sup> Keegan, J. y Lecher, C. (24 de marzo de 2021). *Nevada lawmakers introduce privacy legislation after markup investigation into vaccine websites – the Markup*. The Markup. Consultado a 12 de enero de 2022, disponible en: <https://themarkup.org/blacklight/2021/05/18/nevada-lawmakers-introduce-privacy-legislation-after-markup-investigation-into-vaccine-websites>

El cuidado a la hora de diseñar y programar los algoritmos que alimentan Blacklight y auditan las páginas introducidas por el lector o usuario, se pone de manifiesto al introducir la url de *The Markup* en el propio *Blacklight*.

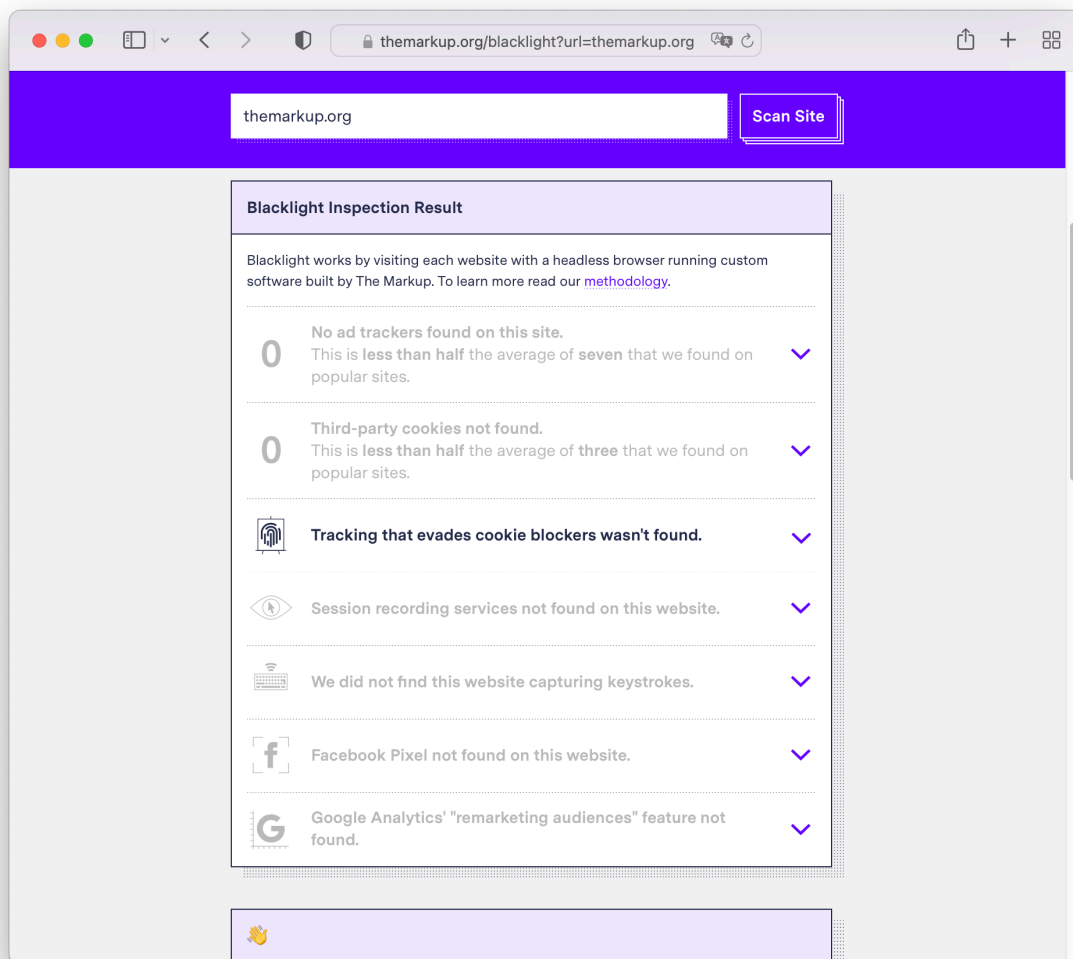
La herramienta, tras analizar la web de *The Markup*, responde con el siguiente mensaje:

“Has solicitado investigar la web de *The Markup*, el medio de comunicación sin ánimo de lucro que creo esta herramienta que estás usando.

*The Markup* no sólo ha valorado la posibilidad de que el lector audite su propia web sino que además su web refleja los valores que proyectan en el periodismo de investigación que llevan a cabo:

“Hola. Es posible que te hayas percatado de que el resultado de que el resultado del análisis es completamente limpio. Esto se debe a que nuestra política de privacidad pretende registrar la mínima información de nuestros lectores que nos sea posible. No utilizamos cookie ni vendemos los datos de nuestros usuarios al mercado de la publicidad. Confía en nosotros, no ha sido fácil construir una página web sin rastreadores digitales. Tu privacidad lo merecía”.

Figura 10. Resultado de la búsqueda de la web de *The Markup* (themarkup.org) en la herramienta *Blacklight*.



Fuente: *The Markup*.

Después de detallar por qué el lector no encontrará ningún rastreador en la web de *The Markup*, *Blacklight* prosigue con una llamada para financiar el tipo de periodismo que realizan:

“Como medio de comunicación sin ánimo de lucro, nos servimos de donaciones para financiar nuestro trabajo. Si esta herramienta te ha sido útil o si quieres ayudarnos a producir más investigaciones para que las grandes compañías tecnológicas rindan cuentas, por favor únete a la tropa de lectores que están financiando nuestra importante misión. Estamos en esto juntos y muy agradecidos por tu apoyo”.

*The Markup* es también un interesante caso de estudio por cómo cuida los detalles en torno a su cobertura y evita cualquier tipo de incongruencia entre sus investigaciones y las prácticas que realiza . Como definen en su política de privacidad <sup>121</sup>, su misión es investigar cómo las instituciones más poderosas del planeta utilizan la tecnología para cambiar la sociedad.

Por ello, y para ser consecuentes con sus métodos e investigaciones, prometen a sus lectores registrar la mínima información personal que les sea posible cuando acudan a la web de *The Markup* para leer sus artículos e investigaciones. En caso de no puedan evitar recopilar esa información, prometen no monetizar esos datos. Además, *The Markup* es un medio que surge con el objetivo de no mostrar anuncios en su web ya que dichos anuncios incluyen a menudo tecnología para rastrear los movimientos del internauta.

---

<sup>121</sup> The Markup. (s.f.). *Privacy policy*. The Markup. Consultado a 12 de enero de 2022, disponible en: <https://themarkup.org/privacy>

Otros proyectos desarrollados en la misma línea por *The Markup* <sup>122</sup> son *Pixel Hunt* <sup>123</sup>, un proyecto que analiza de qué forma la red social Facebook espía y recoge información de sus usuarios, y *Citizen Browser* <sup>124</sup>, una iniciativa que examina qué tipo de contenido promocionan y muestran a sus usuarios las redes sociales como Twitter o Facebook.

## 4.2. Algoritmos al servicio del análisis de datos

Por otro lado, ya sea para mejorar procesos del día como para recopilar datos o investigar otros algoritmos, los profesionales de la información han comenzado a programar y a desarrollar sus propios algoritmos.

Como se ha definido al comienzo de este capítulo, un algoritmo es una serie de reglas que producen un resultado. Por ello, para diseñar, escribir y, en última instancia, crear un algoritmo es necesario conocer un lenguaje de programación y, lo más importante, tener una noción del proceso a ejecutar por el algoritmo y los objetivos de su aplicación.

Es en este preciso momento cuando el profesional de la información utiliza los conocimientos y técnicas de otras disciplinas como las matemáticas, la informática o la estadística. Este conocimiento multidisciplinarios ayudan al periodista o equipo de periodistas a desarrollar los algoritmos que les permiten recopilar, procesar y analizar datos para elaborar informaciones.

---

<sup>122</sup> Fischer, S. (18 de enero de 2022). *The markup doubles down on data tools as revenue grows*. Axios. Consultado a 12 de enero de 2022, disponible en: <https://www.axios.com/the-markup-nonprofit-tech-news-2021-revenue-9bb7a79e-afca-4b30-a791-1deed46b1eb4.html>

<sup>123</sup> The Markup (16 de octubre de 2020). *The citizen browser project-auditing the algorithms of disinformation*. The Markup. Consultado a 14 de enero de 2022, disponible en: <https://themarkup.org/citizen-browser>

<sup>124</sup> Waller, A. y Lecher, C. (21 de enero de 2022). *Help us investigate Facebook Pixel Tracking – the markup*. Help Us Investigate Facebook Pixel Tracking – The Markup. Consultado a 14 de enero de 2022, disponible en: <https://themarkup.org/pixel-hunt/2022/01/21/help-us-investigate-facebook-pixel-tracking>

Python, R y JavaScript son algunos de los lenguajes de programación que facilitan el trabajo del periodista para realizar las labores y acciones anteriores. El periodista debe conocer las sintaxis y gramática de estos lenguajes para poder trabajar con ellos. Su aprendizaje requiere dedicación y horas de práctica.

A pesar de la curva de aprendizaje que supone el uso de estos lenguajes para hacer periodismo, su conocimiento ofrece una serie de ventajas al profesional de la información a la hora de analizar y presentar la información.

Además, el conocimiento de estos lenguajes permite al periodista automatizar ciertos procesos y ejecutar labores intensivas de análisis de datos en un breve período de tiempo, ofreciéndole la posibilidad de reaccionar a la última hora de forma rápida y certera.

#### **4.2.1. Análisis de datos en la investigación *The Opioid Files***

Un ejemplo de los caminos que la automatización, los algoritmos y los lenguajes de programación han abierto para el periodismo es la investigación *The Opioid Files* publicada por *The Washington Post* <sup>125</sup>.

Esta investigación surge tras una batalla legal por obtener los datos de la base de datos *ARCOS (Automation of Reports and Consolidated Orders System)* en propiedad de la Agencia Federal Antinarcoóticos (DEA) y las compañías farmacéuticas. Esta base de datos registra el número de ventas de opioides, la farmacéutica productora del fármaco y el lugar o farmacia de distribución del medicamento.

La batalla legal encabezada por *The Washington Post*, *The Charleston Gazette-Mail of West Virginia* y algunas de las comunidades locales más afectadas por el consumo desenfrenado de opioides derivó en una resolución judicial en junio de 2019 que obligaba a

---

<sup>125</sup> The Washington Post. (20 de Julio de 2019). *Follow the Post's investigation of the opioid epidemic*. The Washington Post. Consultado a 5 de julio de 2022, disponible en: <https://www.washingtonpost.com/national/2019/07/20/opioid-files/>

liberar los documentos que la administración y las compañías farmacéuticas habían mantenido en secreto hasta entonces <sup>126</sup>.

Junto con la pandemia de la COVID-19, la epidemia de opioides es una de las crisis sanitarias más grandes que ha afrontado Estados Unidos en el último medio siglo. Los opioides — su falta de regulación y la prescripción masiva de los mismos — han dejado miles de muertos en Estados Unidos en las últimas décadas. Para poner los números en contexto, el número de fallecidos por sobredosis en 2021 en Estados Unidos sobrepasó la barrera de las 100.000 personas <sup>127 128</sup>.

Con la obtención de los registros de venta y distribución de opioides a nivel nacional, *The Washington Post* pretendía analizar el fenómeno e identificar las comunidades más afectadas por la venta desenfadada de fármacos.

Una vez la decisión judicial fue firme, se estableció un plazo para que la agencia DEA liberase la base de datos ARCOS con los registros de venta de opioides. Este hecho puso en alerta a otros medios de comunicación que por fin tendrían la oportunidad de acceder a los datos. ¿Qué hizo *The Washington Post* para mantener su ventaja en la cobertura de la epidemia de los opioides que había forjado durante años de batallas judiciales? Los reporteros del equipo de Datos del diario buscaron la forma más efectiva para analizar los datos en el menor tiempo posible. Para ello, recurrieron a la automatización.

Para el análisis de esta base de datos, los reporteros de *The Washington Post* se encontraron con una serie de interrogantes y problemas. En primer lugar, tenían la meta de

---

<sup>126</sup> Lentacker, A. (2021). *The Opioid Documents: A Report on the Politics of the Public Record*, *The Social History of Alcohol and Drugs* 2021 35:1, 137-165. Disponible en: <https://www.journals.uchicago.edu/doi/full/10.1086/713409>

<sup>127</sup> CDC. (17 de noviembre de 2021). *Drug overdose deaths in the U.S. top 100,000 annually*. Centers for Disease Control and Prevention. Consultado a 2 de enero de 2022, disponible en: [https://www.cdc.gov/nchs/pressroom/nchs\\_press\\_releases/2021/20211117.htm](https://www.cdc.gov/nchs/pressroom/nchs_press_releases/2021/20211117.htm)

<sup>128</sup> CDC. (11 de mayo de 2022). *Vital statistics rapid release - Provisional drug overdose data*. Centers for Disease Control and Prevention. Consultado a 2 de enero de 2022, disponible en: <https://www.cdc.gov/nchs/nvss/vsrr/drug-overdose-data.htm>

analizar millones de datos en tiempo récord. En segundo lugar, aunque sabían que los datos se liberarían en los próximos días a la resolución judicial, desconocían de antemano el formato y la estructura de los datos. Ese desconocimiento les impedía planear con antelación el proceso para analizar los datos.

Por otro lado, ARCOS es una base de datos masiva con millones de registros. Por tanto, un ordenador de gama media o alta presenta limitaciones para procesar millones de datos en un breve período de tiempo. Además, en caso de procesar los datos, el ordenador puede tardar horas, incluso días en finalizar el proceso.

Una vez finalizado con éxito el procesamiento y análisis de los datos, se presenta el problema de la visualización de miles de elementos en un mapa. En este caso se debe ser selectivo con la tecnología que se usa para esta función ya que no todos los lenguajes de programación soportan el procesamiento de tantos datos, ni la web es capaz de visualizar miles y miles de elementos al mismo tiempo. En cualquier caso, este estudio de caso obvia los elementos relativos a la visualización de los datos, disciplina en la que se profundizará en el próximo capítulo.

Para analizar los datos, *The Washington Post* utilizó el lenguaje de programación R, un entorno y lenguaje de programación abierto y libre que proporciona una variedad de técnicas estadísticas y gráficas.

Más allá del lenguaje utilizado para el análisis de datos, en este caso de estudio destaca la técnica utilizada por los periodistas de datos del diario estadounidense. Con el fin de recortar los tiempos utilizaron una técnica de computación paralela. Esta técnica computacional utiliza dos o más procesadores para realizar diferentes tareas de forma simultánea. Es decir, el ordenador solicita la ayuda de todos sus procesadores para analizar datos de forma simultánea en cada uno de ellos.

Gracias a la planificación previa del algoritmo para analizar los datos y a la técnica a seguir para reducir los tiempos de análisis, los reporteros del diario estadounidense pudieron

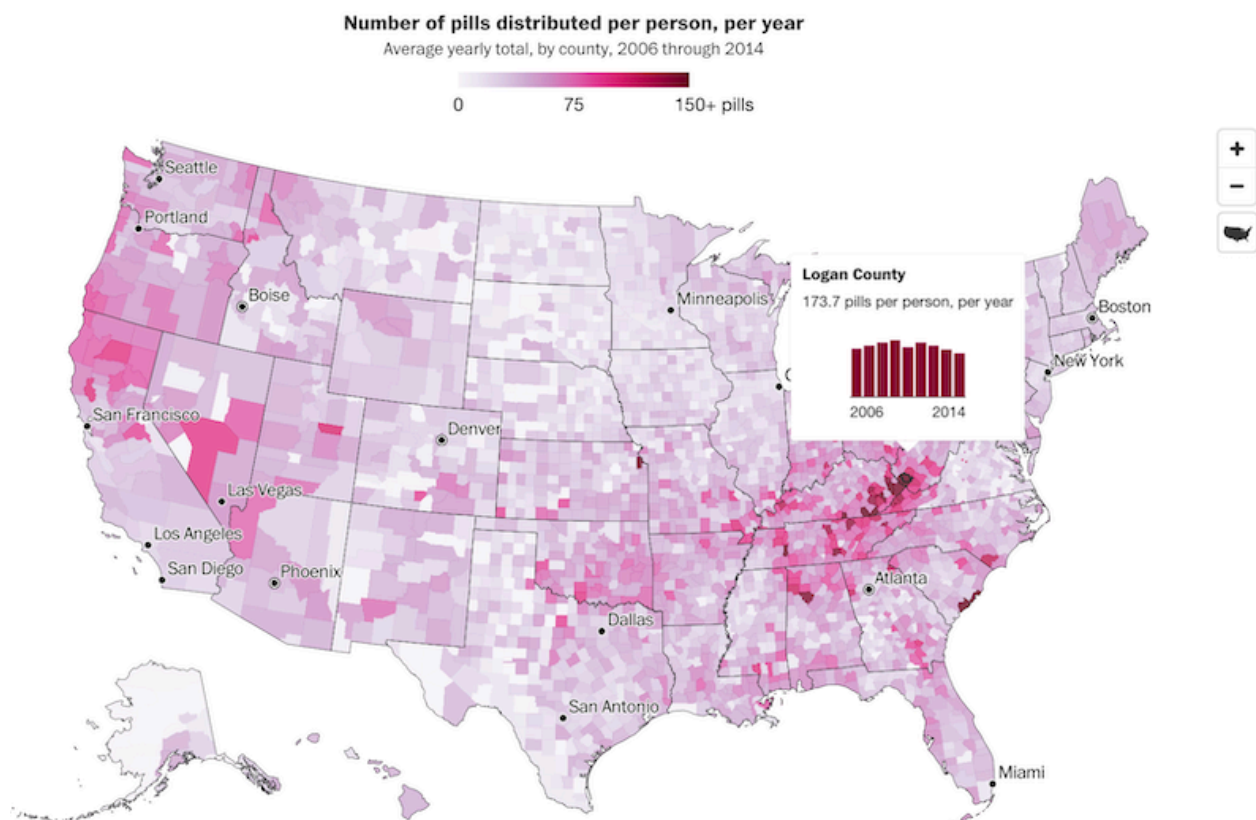
reaccionar con velocidad una vez tuvieron acceso a los datos. El procesamiento y el cuidadoso análisis de los datos derivó en números artículos de investigación que dieron forma a *The Opioid Files*, investigación finalista de los premios Pulitzer.

El resultado final de la investigación y de los procesos de automatización aplicados al periodismo de datos puede verse en el siguiente mapa interactivo que *The Washington Post* incluyó en su cobertura de la epidemia de opioides.<sup>129</sup>

---

<sup>129</sup> The Washington Post. (21 de Julio de 2019). *Drilling into the DEA's Pain Pill Database*. The Washington Post. Consultado a 4 de diciembre de 2020, disponible en: <https://www.washingtonpost.com/graphics/2019/investigations/dea-pain-pill-database/>

Figura 11. Mapa del número de opioides distribuidos por año en cada condado de Estados Unidos.



Fuente: *The Washington Post*.

Además, *The Washington Post* liberó parte del código utilizado en el análisis <sup>130</sup> <sup>131</sup> y creación de visualizaciones una vez publicada la pieza <sup>132</sup>. También puso a disposición del público los datos procesados, estructurados y limpios para que otros medios locales y regionales pudiesen beneficiarse de los mismos y utilizarlos como base de historias locales y regionales <sup>133</sup>.

La velocidad en la publicación de este tema no hubiese sido posible hace unos años sin los avances tecnológicos actuales y ni mucho menos podría haberse ejecutado en tan sólo unos pocos días.

### **4.3. El uso de sensores en el periodismo de investigación**

Sensor journalism o periodismo de sensores es otra de las disciplinas que han surgido en el presente siglo con el objetivo de recabar datos, cuyo análisis informe y alimente reportajes visuales y de investigación. Las bases de datos pueden provenir de un tercero (investigadores, organismos, empresas privadas, entre otros) o pueden ser creadas desde cero por el periodista. Para ello, el profesional de la información puede observar e introducir datos manualmente o servirse de sensores para recoger esos datos.

---

<sup>130</sup> El código y los datos utilizados en la investigación *The Opioid Files* publicada por *The Washington Post* puede consultarse en el siguiente repositorio en GitHub: <https://github.com/wpinvestigative/arcos>

<sup>131</sup> La documentación detallada para el uso del código y los datos de la investigación *The Opioid Files* publicada por *The Washington Post* puede consultarse en el siguiente enlace: <https://wpinvestigative.github.io/arcos/>

<sup>132</sup> Rich, S., Sánchez Díez, M. y Vongkiatkajorn, K. (3 de septiembre de 2019). *How to download and use the DEA pain pills database*. The Washington Post. Consultado a 5 de diciembre de 2020, disponible en: <https://www.washingtonpost.com/national/2019/07/18/how-download-use-dea-pain-pills-database/>

<sup>133</sup> Más información sobre la investigación en la candidatura al galardón *Sigma Awards* de periodismo de datos e investigación: <https://sigmaawards.org/the-opioid-files-2/>

El periodismo de sensores es otra de las disciplinas en que la automatización, la programación, el uso de software y de hardware, en forma de sensores, juega un papel muy relevante.

Para poder analizar el uso de la automatización en esta rama periodística es conveniente fijar primero un marco teórico sobre lo que supone el periodismo de sensores.

Esta disciplina promueve el uso del internet de las cosas, de sensores en definitiva, para recabar información y datos. Estos datos pueden dar pie a revelar informaciones que no se conocían o a trabajar en piezas periodísticas que de otra forma no podrían elaborarse. Es decir, el periodista utiliza los sensores para recopilar y convertir datos en reportajes e investigaciones.

Según BENTON, el periodismo de sensores es la especialidad que busca servirse del internet de las cosas mediante el uso de sensores para recopilar datos que informen las pesquisas de los reporteros o sean la base sobre la que construyan sus historias <sup>134</sup>.

El papel fundamental de los sensores en esta especialización es la capacidad para detectar y almacenar, en forma de datos, características del mundo como pueden ser la luz, la temperatura, el sonido, la presión, la vibración o la calidad del aire, ente otras <sup>135 136</sup>.

---

<sup>134</sup> Benton, J. (4 de agosto de 2016). *Sensor journalism may have lost some of its buzz, but it's also gotten cheaper and easier to pull off*. Nieman Lab. Consultado a 3 de junio de 2019, disponible en: <https://www.niemanlab.org/2021/08/sensor-journalism-may-have-lost-some-of-its-buzz-but-its-also-gotten-cheaper-and-easier-to-pull-off/>

<sup>135</sup> Keefe, J. (2016). *Family Projects for Smart Objects: Tabletop Projects That Respond to Your World*. United States: Make Community, LLC.

<sup>136</sup> Pitt, F. (2014). *Sensors and Journalism* Tow Center for Digital Journalism, A Tow/Knight Report New York, NY, p. 15.

Respecto al uso que hacen los periodistas de los sensores, WEISS destaca tres tipos de utilidades basándose en algunas de las primeras historias que fueron posibles gracias al uso de sensores <sup>137</sup>.

En primer lugar, los periodistas pueden usar sensores para obtener información y datos para probar un suceso a nivel local. Es el caso de *The Houston Chronicle* y su proyecto de calidad del aire publicado en 2005 <sup>138</sup>.

Otro uso consiste en la colaboración de los lectores para investigar un hecho. En este caso, el periodista se sirve de la colaboración de los lectores o usuarios que instalan los sensores en sus casas para construir una base de datos conjunta y representativa sobre un hecho. Un ejemplo es el proyecto desarrollado por la radio neoyorquina WNYC en 2006 para investigar cómo las altas temperaturas afectaban a la salud de los residentes del barrio de Harlem durante el verano. Para hacer posible esta historia, decenas de voluntarios instalaron sensores en sus casas para recoger datos que WNYC analizó más tarde <sup>139</sup>.

Por último, Schmitz Weiss destaca otro uso como es servirse de los datos recogidos por redes de sensores de terceros que ya están en funcionamiento. Es el caso del medio indio *The Hindustan Times* que utilizó sensores gubernamentales, además de sus propios sensores para mediar la calidad del aire en India <sup>140</sup>.

---

<sup>137</sup> Schmitz Weiss, A. (2016). *Sensor journalism: Pitfalls and possibilities*. *Palabra Clave*, 19(4), 1048-1071. Disponible en: [http://www.scielo.org.co/scielo.php?script=sci\\_arttext&pid=S0122-82852016000401048#B14](http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0122-82852016000401048#B14)

<sup>138</sup> Capiello, D. (29 Julio de 2011). *Chronicle cross-county study reveals risky load of 'air toxics'*. *Chron*. Consultado a 6 de junio de 2020, disponible en: <https://www.chron.com/news/article/Chronicle-cross-county-study-reveals-risky-load-1643020.php>

<sup>139</sup> Prasad, S. (7 de Julio de 2016). *This summer, WNYC will investigate New York Heat with sensors and volunteers*. *Poynter*. Consultado a 6 de junio de 2020, disponible en: <https://www.poynter.org/tech-tools/2016/this-summer-wnyc-will-investigate-new-york-heat-with-sensors-and-volunteers/>

<sup>140</sup> Wang, S. (18 de mayo de 2016). *The Hindustan Times is working to build the definitive online source of real-time air quality in all of India*. *Nieman Lab*. Consultado a 6 de junio de 2020, disponible en: <http://www.niemanlab.org/2016/05/the-hindustan-times-is-working-to-build-the-definitive-online-source-of-real-time-air-quality-in-all-of-india/>

En definitiva, el periodismo de sensores ha sido posible gracias a la evolución del internet de las cosas. El desarrollo de chips y ordenadores más pequeños y el abaratamiento de los costes ha propiciado que se popularice el uso de sensores para recopilar datos de los que décadas atrás no se tenían constancia.

Dos de las coberturas donde el periodismo de sensores ha demostrado su gran utilidad es el cambio climático y la contaminación. En los últimos años, se han publicado numerosos proyectos en torno a estos temas usando datos recopilados por sensores y procesos desarrollados por profesionales de la información.

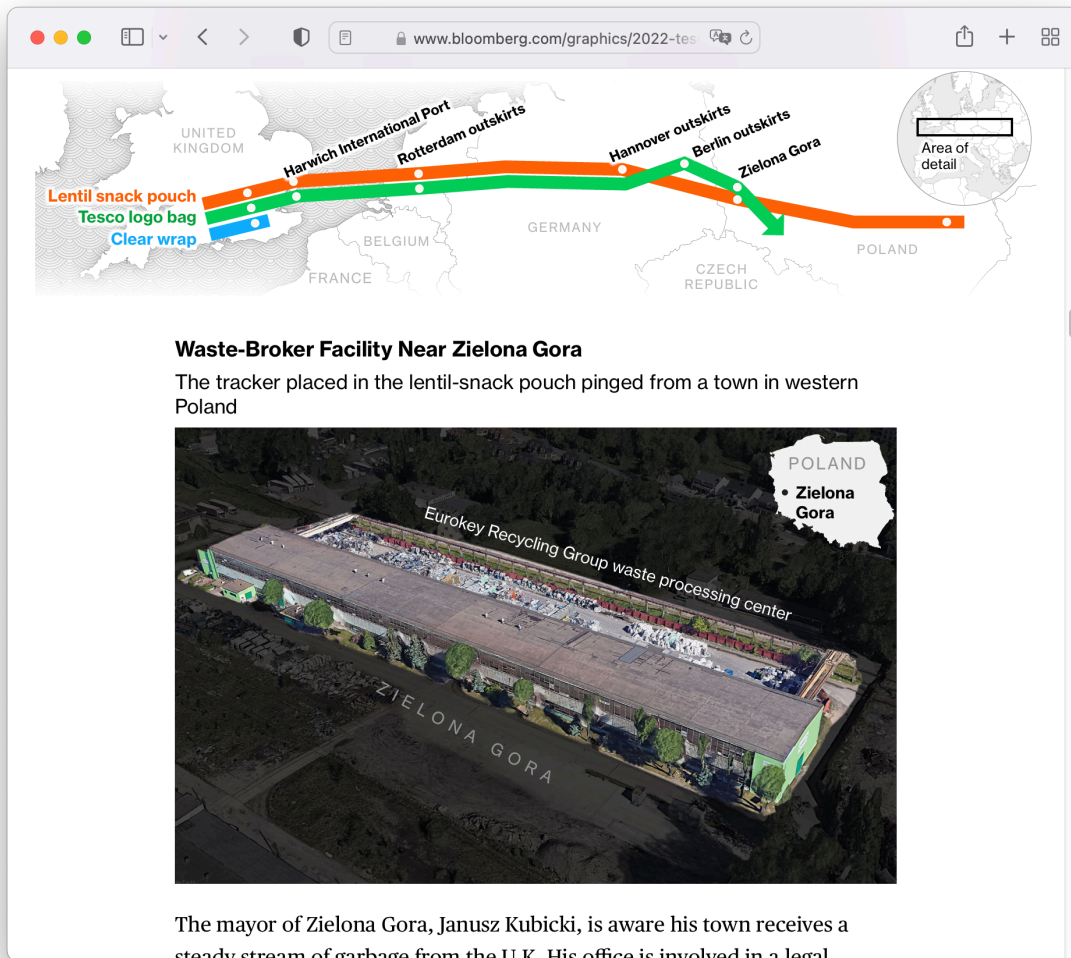
Un ejemplo más reciente es la investigación realizada por Bloomberg <sup>141</sup> en la que sus periodistas realizaron un seguimiento a tres envoltorios de plástico desde el momento en que fueron depositados en un contenedor de plástico en Londres hasta que llegaron al punto de reciclaje.

Ante la imposibilidad de seguir físicamente el recorrido de una bolsa en la cadena de reciclaje, los periodistas de Bloomberg introdujeron un sensor en cada una de esos envoltorios de plástico. Dichos sensores permitieron dibujar el recorrido que normalmente sigue un envoltorio tras ser depositado en el contenedor de reciclaje.

---

<sup>141</sup> Chellel, K., Mokwa, W., Lin, J. C. F. y Randolph, S. (29 de marzo de 2022). *A Plastic Bag's 2,000-Mile Journey Shows the Messy Truth About Recycling*. Bloomberg. Consultado a 12 de abril de 2022, disponible en: <https://www.bloomberg.com/graphics/2022-tesco-recycle-plastic-waste-pledge-falls-short/>

Figura 12. Extracto del artículo *A Plastic Bag's 2,000-Mile Journey Shows the Messy Truth About Recycling* publicado por *Bloomberg* que muestra el recorrido de las tres bolsas de plástico rastreadas mediante el uso de sensores diminutos.



Fuente: Bloomberg.

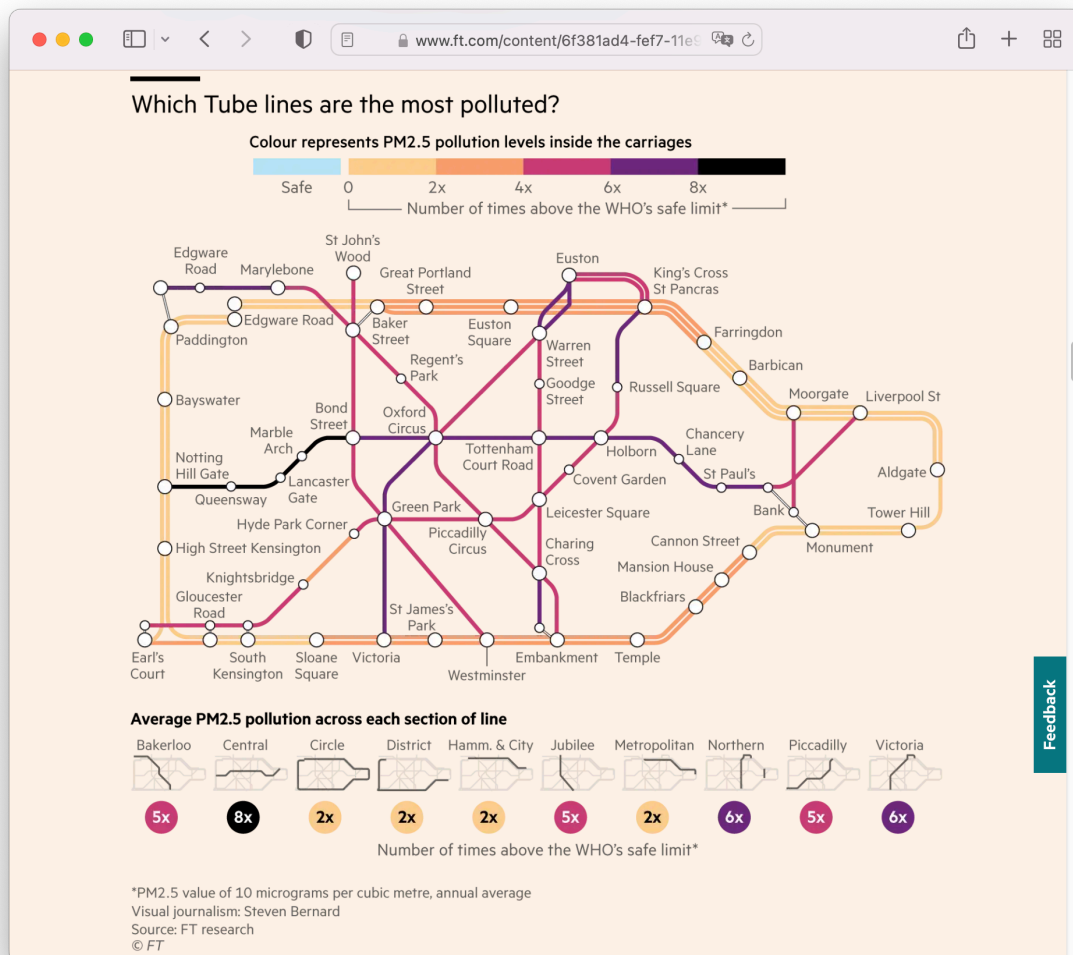
En noviembre de 2019, Financial Times publicó una investigación <sup>142</sup> en la que sus periodistas se sirvieron de sensores para determinar la calidad del aire que se respira en el metro de Londres.

Para medir la calidad del aire en el metro de Londres, los reporteros de Financial Times recurrieron a dos dispositivos portátiles para rastrear los niveles de contaminación dentro de los vagones de las líneas que recorren la zona centro de Londres. Estos dispositivos, fabricados por las compañías Plantower y Plume Labs, registraron la presencia en el aire de partículas consideradas peligrosas para la salud humana.

---

<sup>142</sup> Hodgson, C., Hook, L. y Bernard, S. (5 de noviembre de 2019). *London Underground: The dirtiest place in the city*. Financial Times. Consultado a 10 de abril de 2022, disponible en: <https://www.ft.com/content/6f381ad4-fef7-11e9-be59-e49b2a136b8d>

Figura 13. Diagrama del mapa de metro de Londres publicado por Financial Times con los niveles de contaminación por tramos de vía.



Fuente: Financial Times.

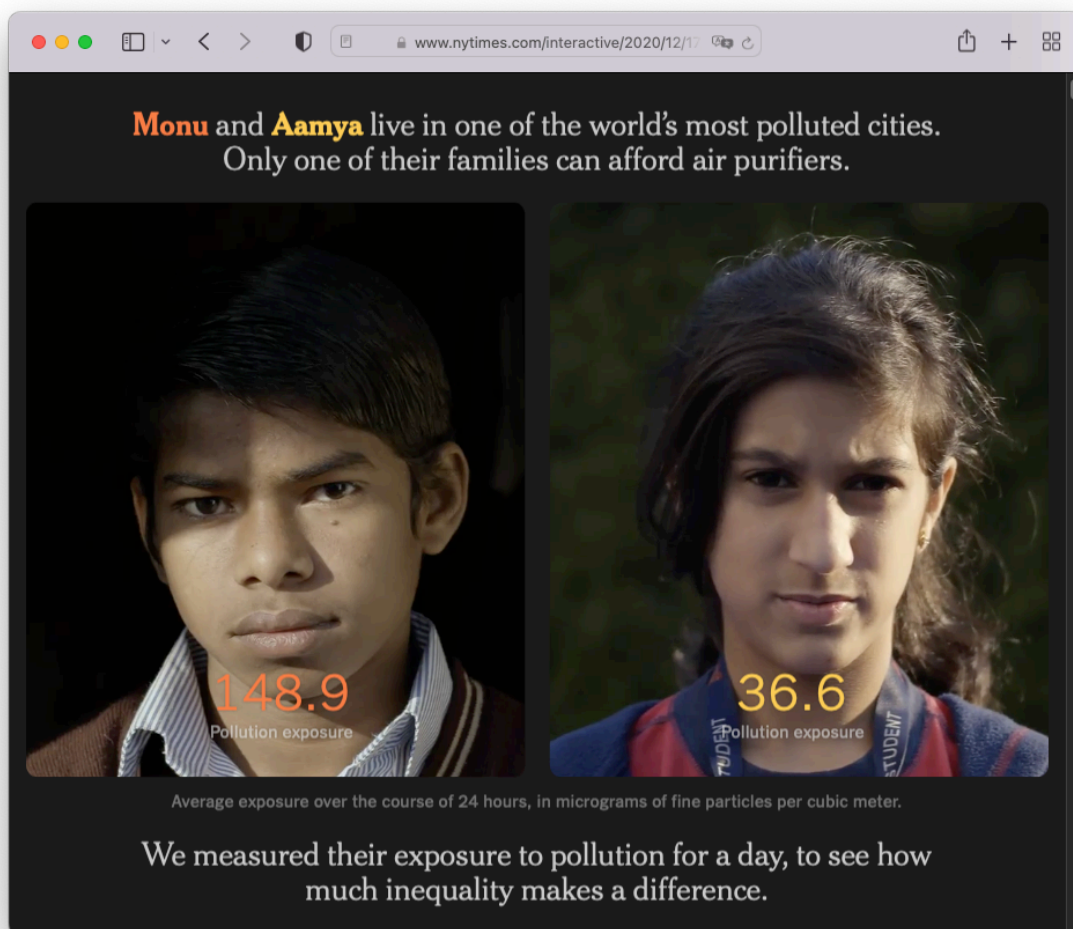
A nivel urbano y local, *The New York Times* publicó una historia visual titulada *Who Gets to Breathe Clean Air in New Delhi?* sobre las diferencias a nivel de contaminación y aire respirado en un barrio rico y en un barrio pobre en Nueva Delhi, India <sup>143</sup>.

Mediante el uso de sensores, los periodistas de *The New York Times* midieron durante un día la calidad del aire a la que estaban expuestos dos niños, Monu y Aamyra, pertenecientes a una familia sin recursos y a una familia acaudalada respectivamente. De esta forma, pudieron mostrar con datos y gráficos cómo la contaminación tiene un mayor impacto en las familias con menos recursos.

---

<sup>143</sup> Wu, J., Watkins, D., Williams, J., Bhagat, S. V., Kumar, H., Gettleman, J., Taylor, R., Davis, L. y Singh, K. D. (17 de diciembre de 2020). *Who gets to breathe clean air in New Delhi?* *The New York Times*. Consultado a 10 de abril de 2022, disponible en: <https://www.nytimes.com/interactive/2020/12/17/world/asia/india-pollution-inequality.html>

Figura 14. Introducción del artículo *Who Gets to Breathe Clean Air in New Delhi?* publicado por *The New York Times*.



Fuente: *The New York Times*.

En definitiva, los sensores otorgan al periodista la capacidad de capturar datos y crear nuevas bases de datos que puede explotar y analizar en sus historias. En particular, los profesionales de la información han sabido sacar el máximo partido de los sensores en la cobertura del clima, del medio ambiente y de la salud pública debido a facilidad de los sensores para registrar diferentes tipos de datos biológicos físicos o sociales <sup>144</sup>.

A pesar del potencial que tienen los sensores para la recogida de datos y la identificación de nuevas, la ejecución de sus funciones debe ser tomada con cautela y debe ser supervisada por un periodista para evitar cualquier error en la recolección y análisis de los datos <sup>145</sup>.

Con el objetivo de enfatizar la novedad que supone este género, BENTON hace una reflexión interesante:

“Here’s to “sensor-assisted reporting” someday sounding as outdated as “computer-assisted reporting,” and to this sort of work someday being a standard part of a reporter’s toolkit.”

Benton espera que, algún día, el periodismo de sensores suene tan anticuado como el periodismo asistido por ordenador. Según el autor, eso querrá decir que los sensores ya forman parte de las herramientas de trabajo del periodista.

---

<sup>144</sup> Fahn, J. (s.f.). *Feel me?*, Columbia Journalism Review. Consultado a 5 de abril de 2022, disponible en: [https://archives.cjr.org/feature/feel\\_me.php](https://archives.cjr.org/feature/feel_me.php)

<sup>145</sup> Schmitz Weiss, A. (2016). Sensor journalism: Pitfalls and possibilities. *Palabra Clave*, 19(4), 1048-1071.

#### 4.4. Machine learning aplicado al periodismo de investigación

Otra alternativa para el periodismo de investigación es aplicar sistemas de automatización inteligentes o inteligencia artificial. El machine learning es una de las técnicas de inteligencia artificial que más se han desarrollado a nivel computacional en los últimos años. Como se ha visto en el Capítulo II, el desarrollo ha sido tal que los periodistas han hecho uso del mismo para realizar investigaciones que de otra forma no se podrían realizar.

Estos algoritmos inteligentes permiten tomar decisiones complejas de forma rápida y escalable – es decir son capaces de analizar grandes conjuntos de datos – mediante la automatización. De acuerdo con Diakopoulos, las cuatro decisiones o acciones juiciosas más comunes que un algoritmo al servicio del periodismo puede tomar son <sup>146</sup>:

- Priorizar
- Clasificar
- Asociar
- Filtrar

Veámoslo con un ejemplo. Un caso de estudio sobre cómo la inteligencia artificial puede dotar de nuevos y poderosos recursos a los periodistas para mejorar sus investigaciones es la serie de historias titulada *Doctors and Sex Abuse* (Médicos y abuso sexual, por su traducción al español) publicada en 2016 por *The Atlanta Journal-Constitution* <sup>147</sup>. Como se expone a continuación, los periodistas de *The Atlanta Journal-Constitution* utilizaron un algoritmo de clasificación basado en machine learning para clasificar informes y datos.

---

<sup>146</sup> Diakopoulos, N. (2019). *Automating the News: How Algorithms Are Rewriting the Media*. United Kingdom: Harvard University Press, pp. 19-22.

<sup>147</sup> The Atlanta Journal-Constitution. (s.f.). *Doctors & Sex Abuse*. The Atlanta Journal-Constitution. Consultado a 21 de noviembre de 2022, disponible en: <https://doctors.ajc.com/>

La investigación expone cientos de casos de abuso sexual entre doctores estadounidenses y sus pacientes. Además, la serie de artículos de *The Atlanta Journal-Constitution* profundiza en la ausencia de consecuencias que en muchos casos han tenido estas acciones.

La investigación, finalista de los premios Pulitzer en 2017 en la categoría de cobertura nacional, es el paradigma perfecto de cómo la inteligencia artificial puede elevar los contenidos y los reportajes periodísticos. De hecho, esta historia fue posible debido a las técnicas de inteligencia artificial que los reporteros de *The Atlanta Journal-Constitution* utilizaron para analizar miles de documentos. De no haber sido por este análisis automatizado, el reportaje no hubiese podido llevarse a cabo.

En primer lugar, los periodistas de *The Atlanta Journal-Constitution* descargaron mediante una técnica de scraping más de 100.000 documentos que detallaban las sanciones disciplinarias que se habían aplicado a miles de médicos estadounidenses <sup>148</sup>. Para ello, los periodistas tuvieron que adaptar su robot de extracción de datos a las 50 agencias estatales estadounidenses que poseían la información ya que los documentos no estaban disponibles de forma centralizada.

Analizar 100.000 documentos era tarea imposible por lo que los periodistas del medio estadounidense crearon un programa o algoritmo inteligente de machine learning para analizar y clasificar cada caso. En función de palabras clave, el algoritmo otorga a cada documento una puntuación con la probabilidad de que la sanción esté relacionada con un caso de conducta sexual inapropiada del médico.

De esta forma, pudieron reducir el número de documentos a sólo 6,000 casos que analizaron y estudiaron para realizar la investigación.

---

<sup>148</sup> The Atlanta Journal-Constitution. (2016). *About the AJC's investigation of doctor misconduct*. Doctors y Sex Abuse. The Atlanta-Journal Constitution. Consultado a 21 de noviembre de 2022, disponible en: [https://doctors.ajc.com/about\\_this\\_investigation/](https://doctors.ajc.com/about_this_investigation/)



**CAPÍTULO V:**  
**Automatización en el periodismo**  
**visual**



El periodismo robot y la automatización de procesos también forman parte de las visualizaciones, las infografías y los artículos visuales e inmersivos que los departamentos de Datos y Gráficos de los medios de comunicación elaboran en su día a día. Gracias a la automatización, las aplicaciones algorítmicas y otras tecnologías, el periodismo visual ha crecido exponencialmente en los medios de comunicación desde la segunda década del presente siglo <sup>149</sup>.

Las posibilidades que los robots y los lenguajes de programación ofrecen a los profesionales del apartado visual de los medios son inmensas. Gracias a estas tecnologías, los periodistas pueden visualizar datos en segundos o crear experiencias interactivas y visuales para el lector.

---

<sup>149</sup> Carrasco Polaino, R. y Viviar Flores, J. (2020). *Periodismo de datos y visualización en Periodismo y nuevos medios: Perspectivas y retos*, pp. 249–273. Editorial GEDISA.

Desde que en 1981, *The Columbus Dispatch*<sup>150</sup> <sup>151</sup> se convirtiese en el primer periódico publicado en internet mediante el servicio CompuServe, la web ha ido evolucionando y ha permitido ofrecer nuevos contenidos a los medios. Cuatro décadas más tarde, las posibilidades que la web y los lenguajes de programación ofrecen, especialmente JavaScript, han provocado la evolución del periodismo visual en forma de piezas interactivas, automatizadas e inmersivas para el lector.

El periodismo visual, en concreto en su rama de gráficos, es uno de los mayores exponentes de cómo la tecnología, la automatización de procesos y los lenguajes de programación pueden ayudar al profesional de la información a contar historias, generar gráficos y explicar contenidos complejos de manera clara y precisa.

El periodista actual que desarrolla su trabajo en un equipo de gráficos posee los conocimientos para crear visualizaciones complejas y atractivas que presentan la información de manera más amena y digerible para el lector. Además, gracias a la evolución de la tecnología y la incorporación de la misma al sector del periodismo, el profesional de la información tiene a su disposición un sinfín de softwares, lenguajes de programación y secuencias de código para automatizar las tareas que realiza en el día a día. Estas nuevas técnicas le permiten realizar tareas complejas y crear elaboradas visualizaciones basadas en miles o millones de datos en tan sólo unas horas.

El periodismo robot y los procesos de automatización están presentes en múltiples tareas necesarias para la elaboración de una infografía o de una pieza visual. Como se verá en

---

<sup>150</sup> Shedden, D. (2 de mayo de 2013). *New media timeline (1980)*. Poynter. Consultado a 6 de febrero de 2021, disponible en: <https://www.poynter.org/archive/2004/new-media-timeline-1980/#:~:text=The%20first%20newspaper%20to%20go,The%20Middlesex%20News%2C%20the%20St.>

<sup>151</sup> Shedden, D. (24 de noviembre de 2014). *Today in media history: CompuServe and the First Online Newspapers*. Poynter. Consultado a 6 de febrero de 2021, disponible en: <https://www.poynter.org/reporting-editing/2014/today-in-media-history-compuserve-and-the-first-online-newspapers/>

las siguientes páginas, algunos de los casos donde la tecnología tiene un papel fundamental en el periodismo visual siguen a continuación:

- El uso de los lenguajes de programación para la creación y producción de gráficos estáticos e interactivos en breves periodos de tiempo posibilitando la reacción a las noticias de última hora.
- El uso de librerías de código propias o de terceros para determinar y crear los diferentes elementos con los que cuenta un gráfico como son, entre otros, los colores, las escalas, el procesamiento y cálculo de datos.
- El uso de herramientas o programas de software propios o de terceros como *Datawrapper*, *Flourish* o *RAWGraphs*, por poner algunos ejemplos, para la creación de gráficos tanto estéticos como interactivos.
- La creación de historias visuales basadas en datos como trackers, simulaciones o *newsgames* cuya actualización es automática o semiautomática debido al carácter variable de la información. La publicación de este tipo de historias visuales suele automatizarse cuando se prevé que la información y los datos van a actualizarse o cambiar de forma periódica. Como se analiza en esta disertación, un ejemplo de este tipo de proyectos visuales son los paneles de datos o trackers que han publicado los medios de comunicación con información actualizada de la pandemia a diario.

## **5.1. El papel de los lenguajes de programación en el periodismo visual**

Los periodistas utilizan los lenguajes de programación para traducir datos complejos en gráficos, abstracciones o formas geométricas a las que la sociedad ha dotado de un sentido con el objetivo de comunicar las conclusiones de un análisis de datos o una historia de forma visual y simplificada.

Las infografías, mapas, gráficos y visualizaciones de datos propias del periodismo visual, que consisten en la exposición gráfica de diferentes elementos con el objetivo de representar y transmitir información, son muy útiles ya que permiten revelar tendencias y patrones bajo montañas de datos y hechos <sup>152</sup>.

Aunque normalmente los términos de infografía o visualización de datos se utilizan de forma intercambiable <sup>153</sup>, lo cierto es que la literatura tiende a marcar ciertas diferencias entre ambas.

Las infografías normalmente están estructuradas de forma lineal, es decir las explicaciones se ordenan de en función del tiempo o la relevancia como se haría en una narrativa clásica o periodística con el formato de la pirámide invertida. Las infografías son ricas en detalles, coloridas y abundantes en elementos como dibujos, iconos o pictogramas. Su objetivo es informar al público de forma clara y comunicar la mayor cantidad de información posible en un mensaje previamente determinado por el periodista, infografista o diseñador.

Por su parte, las visualizaciones de datos presentan un carácter más abierto, dando pie al lector a analizar, explorar y descubrir el gráfico, y los datos detrás de éste, por su cuenta. En lugar de portar un mensaje visual para el lector, el profesional de la información da pie al propio usuario a obtener sus propias conclusiones. Por su carácter exploratorio a veces se consideran herramientas para que el lector aprenda o se informe más sobre un tema <sup>154</sup>. En una visualización de datos, el periodista corre el riesgo de que el mensaje que quiere

---

<sup>152</sup> Cairo, Alberto (2016). *The Truthful Art: Data, Charts, and Maps for Communication*. Berkeley, CA: Peachpit Press.

<sup>153</sup> Pérez-Montoro, M. (2022). *Comunicación visual de la información: qué y cómo podemos narrar con datos*, Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT). Disponible en: [http://diposit.ub.edu/dspace/bitstream/2445/183791/4/P%C3%A9rez-Montoro%20%282022%29%20Comunicaci%C3%B3n%20visual%20de%20la%20informaci%C3%B3n\\_def.pdf](http://diposit.ub.edu/dspace/bitstream/2445/183791/4/P%C3%A9rez-Montoro%20%282022%29%20Comunicaci%C3%B3n%20visual%20de%20la%20informaci%C3%B3n_def.pdf)

<sup>154</sup> Cairo, Alberto (2016). *The Truthful Art: Data, Charts, and Maps for Communication*. Berkeley, CA: Peachpit Press, p. 52.

transmitir en su pieza se difumine debido a la libertad que el lector tiene para leer la visualización e interpretarla.

En cualquier caso, en ambos casos, la tecnología en forma de softwares especializados o lenguajes de programación juega un papel fundamental para facilitar la creación de las infografías o las visualizaciones de datos, automatizando parte del proceso de diseño y creación.

Como se ha descrito en el Capítulo II. Periodismo robot, los lenguajes de programación más comunes entre los profesionales de la información son JavaScript, R y Python. Estos lenguajes de programación, junto a una serie de librerías de código asociadas que facilitan una serie de acciones concretas, son también los lenguajes más populares para la creación de visualizaciones de datos entre los periodistas y diseñadores.

Tabla 2. Lenguajes de programación más comunes para la visualización de datos en los medios de comunicación.

<b>Lenguaje de programación</b>	<b>Librería de código</b>
<i>JavaScript</i>	<i>Canvas</i>
	<i>d3.js</i>
	<i>Leaflet</i>
	<i>Mapbox</i>
	<i>p5.js</i>
<i>Python</i>	<i>Matplotlib</i>
	<i>Seaborn</i>
<i>R</i>	<i>ggplot2</i>

Fuente: Elaboración propia.

Estos lenguajes de programación permiten al periodista traducir un conjunto de datos en una visualización que pretende ofrecer información adicional a un artículo o, en ocasiones, ser el artículo en sí mismo.

Estos lenguajes son muy extensos y están compuestos de diferentes librerías de código o frameworks, archivos o conjuntos de código que facilitan una serie de acciones o funcionalidades. A día de hoy, la mayoría de los gráficos y mapas publicados por medios de comunicación han sido realizados gracias al soporte y la ayuda de diferentes procesos ejecutados por el código o los robots que forman parte de estas librerías de programación.

En Python, uno de los lenguajes más legibles, versátiles por el número de aplicaciones y populares entre programadores <sup>155</sup>, destacan las librerías seaborn y matplotlib como los principales exponentes para la creación de gráficos sencillos. La versatilidad de Python le convierte en un lenguaje navaja en la ciencia de datos y, por ende, en el periodismo debido a que puede utilizarse tanto para analizar como para visualizar grandes bases de datos <sup>156</sup>. Por su parte, ggplot2 es la librería más utilizada para la visualización de datos en R y una de las librerías más populares del propio lenguaje.

Aunque ambos lenguajes permiten crear visualizaciones interactivas, su uso en el sector del periodismo está más enfocado al análisis de datos y a la creación de gráficos estáticos. Los gráficos producidos tanto en R como en Python son base en muchas ocasiones de infografías más complejas que se completan con el software de edición vectorial Adobe Illustrator.

Por su parte, la librería de código de JavaScript se ha convertido en el lenguaje estándar para visualizar datos en la web. El periodista puede realizar gráficos complejos basados en cientos o miles de registros gracias a este lenguaje. JavaScript cuenta con diferentes librerías

---

<sup>155</sup> Tagliaferri, L., Tagliaferri, L. (2018). How To Code in Python 3. United States: DigitalOcean., p.7.

<sup>156</sup> VanderPlas, J. (2016). Python Data Science Handbook: Essential Tools for Working with Data. United States: O'Reilly Media, p.13.

o conjuntos de código y funciones que han sido diseñados con unos objetivos muy concretos. d3.js es una de esas librerías. En los últimos años, d3.js se ha convertido en la librería por excelencia para realizar gráficos y mapas entre los periodistas especializados en la visualización de datos.

d3.js es una librería de código de JavaScript creada por el ex periodista de *The New York Times*, Mike Bostock, en 2011. d3.js facilita la manipulación de miles de datos para, mediante el uso de diferentes funciones, convertirlos en abstracciones de la realidad, reducirlos a figuras geométricas y escalas, produciendo visualizaciones de datos.

Este tipo de visualizaciones complejas e interactivas producidas gracias a d3.js u otras librerías son fundamentales a día de hoy en los medios de comunicación. Las visualizaciones de datos bien diseñadas y ejecutadas facilitan la comprensión por parte del lector de temas complejos y de artículos en los que se analizan cantidades inmensas de datos que no pueden ser procesadas de manera individual.

La estructura de la web, donde el 98% de los portales <sup>157</sup> y páginas online están construidas utilizando JavaScript de una forma u otra en su esqueleto, existe una amplia gama de librerías para visualizar datos en Internet.

Una de las técnicas más utilizadas para la automatización de gráficos y visualizaciones es el uso de Application programming interfaces o interfaces de programación de aplicaciones (APIs). Las APIs son protocolos para acceder a bases de datos de forma organizada, remota y escalable. Es decir, mediante el uso de APIs, el profesional de la información puede “conectar” la visualización final a una base de datos. De esta forma, cada vez que los datos cambian no es necesario actualizar el gráfico. Por tanto, una vez programado el gráfico, el periodista no tendrá que preocuparse en invertir tiempo para actualizar los datos.

---

<sup>157</sup> W3Techs. (s.f.). *Usage statistics of JavaScript as client-side programming language on websites*. W3Techs. Consultado a 8 de enero de 2022, disponible en: <https://w3techs.com/technologies/details/cp-javascript/>

## 5.2. Herramientas automatizadas para la creación de visualizaciones

En el lado del software, los periodistas utilizan diferentes programas para crear tanto gráficos como mapas. QGIS es el software por excelencia para generar mapas<sup>158</sup>. Blender es uno de los programas que más aceptación tiene en la generación de gráficos en 3-D. Adobe Illustrator y Adobe Photoshop son las herramientas preferidas para la edición y creación de gráficos estáticos.

Tabla 3. Aplicaciones de algunas de las herramientas más populares para la creación de visualizaciones de datos en el sector del periodismo.

	Gráficos	Mapas
Adobe Illustrator		
Adobe Photoshop		
Blender		
QGIS		
Datawrapper		
Flourish		
RAWGraphs		
Tableau		

Fuente: Elaboración propia.

Por otro lado, las herramientas de creación de visualizaciones de terceros como Datawrapper, Flourish o RAWGraphs, entre otras, son algunas de las herramientas más populares en las redacciones de los medios de comunicación en la actualidad.

Estas herramientas facilitan la traducción de los datos a elementos gráficos o mapas. Además, proporcionan una solución a los periodistas, desde los especialistas en gráficos a los

---

<sup>158</sup> Olaya, V. (2014). *Sistemas e Información geográfica*. Disponible en: [https://www.icog.es/TyT/files/Libro\\_SIG.pdf](https://www.icog.es/TyT/files/Libro_SIG.pdf)

profesionales sin ninguna experiencia o conocimiento en la disciplina, para crear visualizaciones de datos en apenas segundos.

El uso de herramientas como Datawrapper o Flourish supone la automatización de una parte importante de la visualización de datos o gráficos que produce un medio de comunicación. Por ello, múltiples redacciones han incorporado estas herramientas a sus procesos de trabajo. Aunque el periodista, infografista o diseñador todavía deber realizar parte del trabajo de preparación e introducción de datos y selección del gráfico que desea visualizar, estas herramientas facilitan la producción de la visualización en apenas minutos.

Para conocer de primera mano el papel que juega la automatización en estas herramientas y en la visualización de datos, se analiza a continuación el funcionamiento de Datawrapper. Esta herramienta, que cuenta con una versión gratuita y otra de pago, ha reducido los tiempos y facilitado la creación de gráficos y mapas en las redacciones. Tal es así que lo que comenzó como un experimento, hoy se ha convertido en un negocio que proporciona servicios de visualización a medios como *The New York Times*, *Associated Press* o *Thomson Reuters*.

Mirko Lorenz, fundador de *Datawrapper*, reflexiona sobre los inicios y sobre la evolución de la herramienta que en diez años ha pasado de ser una idea a una de las herramientas más utilizadas por los medios de comunicación.

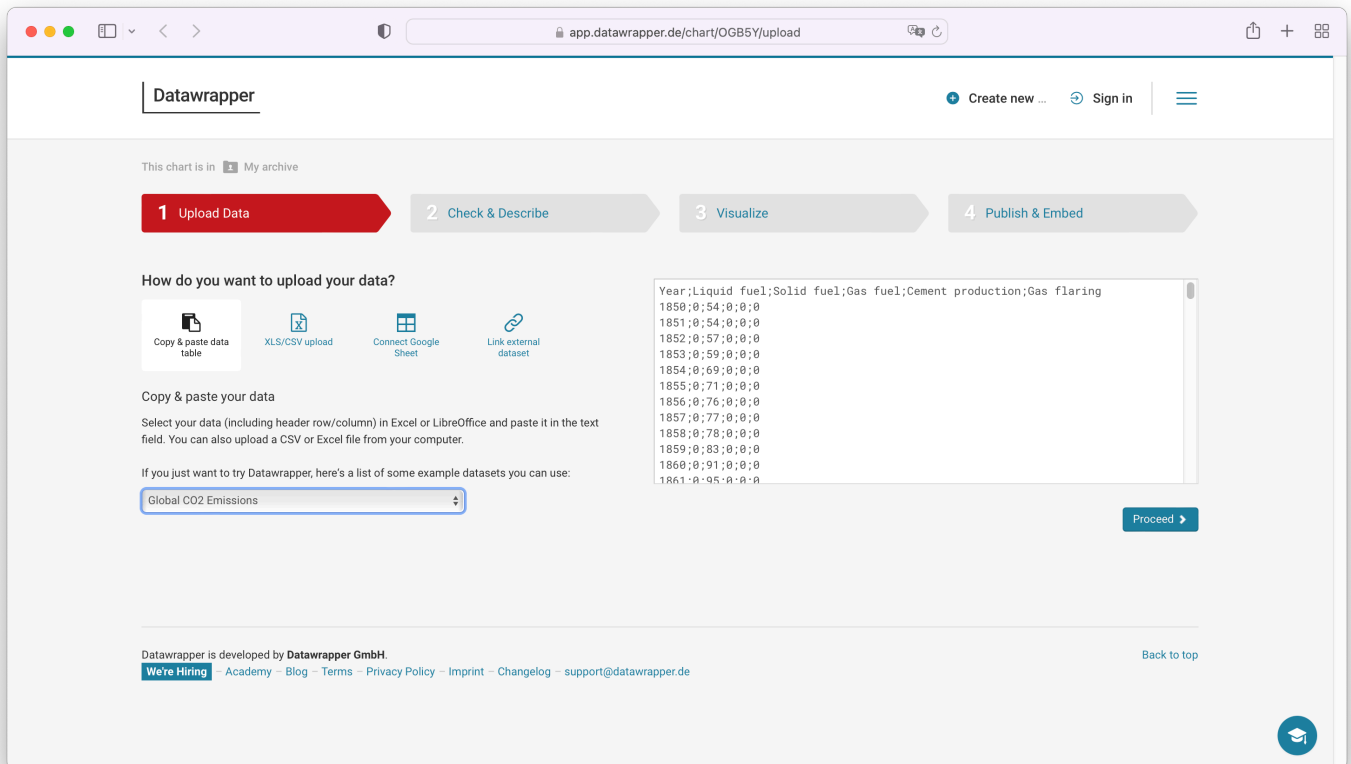
“I like to innovate, to think about how something could be done differently and better. I don’t care about how and why things are the way they are, if there is a better way. In journalism, the use of data was too complicated in 2012. Creating one chart, even just a simple line or bar chart, would mean hours of work for the graphics department. Hours!”<sup>159</sup>

---

<sup>159</sup> Lorenz, M. (23 de febrero de 2022). *Celebrating ten years since the datawrapper beta*. Datawrapper Blog. Consultado a 25 de febrero de 2022, disponible en: <https://blog.datawrapper.de/datawrapper-beta-anniversary/>

El funcionamiento de la herramienta es muy sencillo. En una primera pantalla, el profesional de la información debe introducir los datos que quiere visualizar.

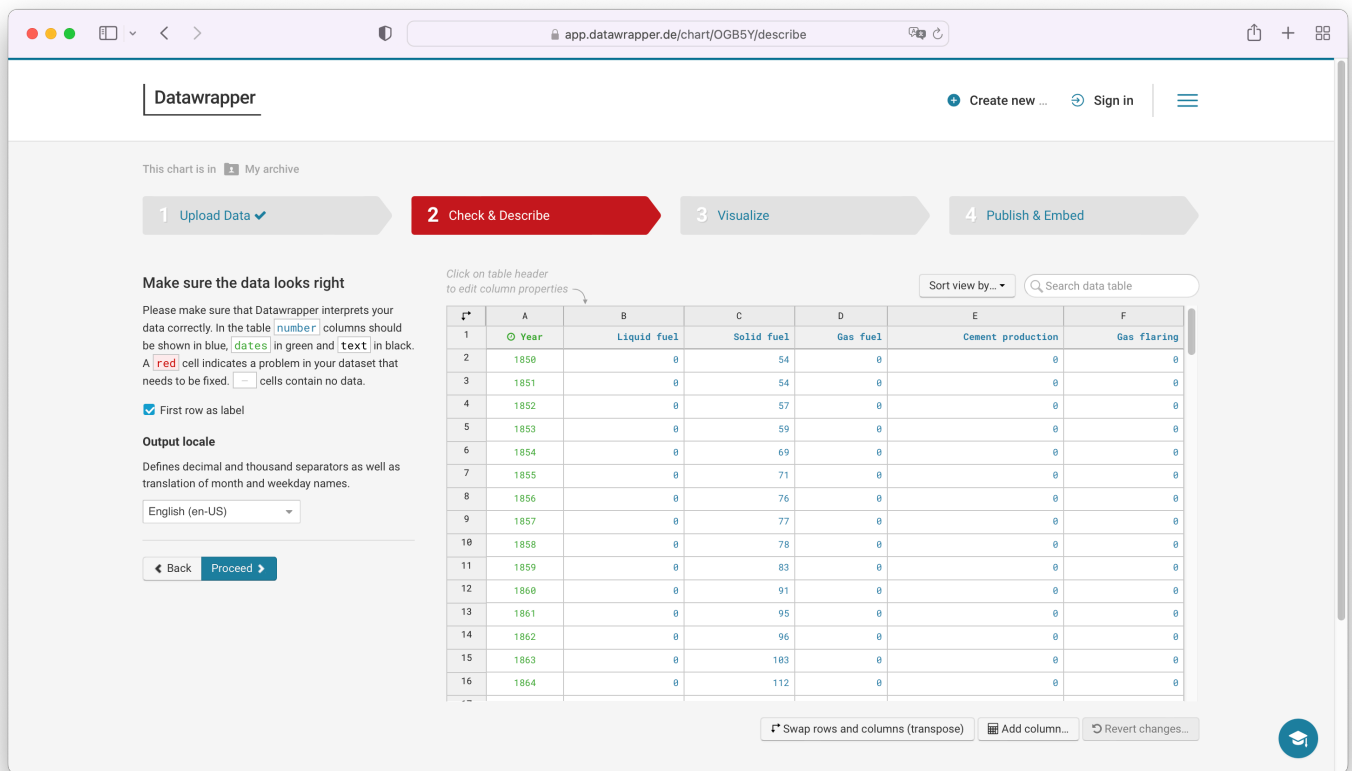
Figura 15. Proceso de subida de datos a la herramienta de creación de gráficos Datawrapper.



Fuente: Datawrapper.

En segundo lugar, el periodista comprueba los datos introducidos y determina qué tipo de dato se corresponde con cada una de las columnas. Además, Datawrapper ofrece la posibilidad de realizar una serie de operaciones con los datos introducidos en caso de que se quiera calcular algún valor adicional. Ese valor se añadirá como una columna extra a los datos estructurados en una tabla en pantalla.

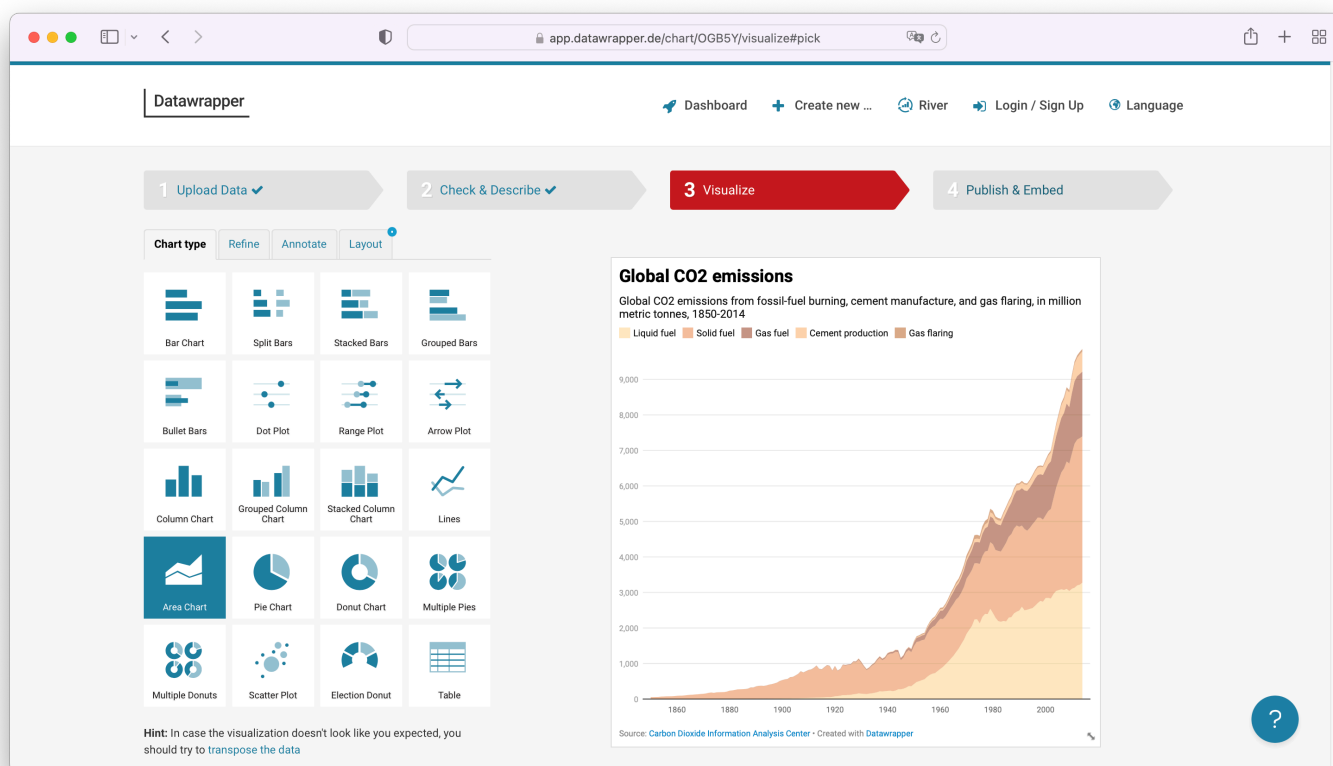
Figura 16. Proceso de chequeo y edición de datos de la herramienta de creación de gráficos Datawrapper.



Fuente: Datawrapper.

Por último, los algoritmos que conforman Datawrapper ofrecen al usuario diferentes posibilidades para visualizar sus datos. El profesional de la información puede elegir hasta veinte tipos de gráfico diferentes para visualizar los datos de forma automática. Además, en apenas unos segundos, puede configurar el aspecto, los colores y los textos presentes en el gráfico.

Figura 17. Proceso de selección de gráficas y visualizaciones de datos de la herramienta de creación de gráficos Datawrapper.



Fuente: Datawrapper.

Una vez seleccionado el gráfico, el periodista tan solo tiene que seleccionar y el gráfico estará disponible en diferentes formatos para ser publicado en la web de su medio de comunicación. Datawrapper también ofrece el mismo proceso para la creación de mapas mediante el uso de datos espaciales.

De esta forma, el profesional de la información puede crear gráficos sencillos en apenas unos pocos segundos. Por un lado, la automatización de procesos presente en las herramientas gráficas como Datawrapper, libera de responsabilidades menores a los especialistas en visualización de datos para que puedan centrarse en historias visuales, gráficos y mapas más ambiciosos.

Por otro lado, la curva de aprendizaje para los reporteros de otras secciones es mínima. Tras retener una serie de conceptos básicos, cualquier reportero puede crear sus propios gráficos sencillos en apenas unos minutos. De esta forma, la aplicación de la automatización en la creación de gráficos ha dado el poder al reportero de crear sus propios gráficos para contextualizar sus historias.

### **5.3. Automatización y periodismo visual en tiempos de pandemia <sup>160</sup>**

Un ejemplo reciente de visualizaciones de datos más elaboradas son las piezas visuales publicadas para explicar el avance y los efectos de la pandemia global de la COVID-19. Las visualizaciones sobre la evolución del virus son una de las principales informaciones que han consumido los lectores para informarse del avance de la pandemia.

Este tipo de información visual sobre la COVID-19 comenzó a elaborarse de forma manual, mediante el uso de infografías, durante las primeras semanas de 2020 cuando la cobertura no había comenzado una dimensión global. Con el paso de las días y según se extendía el virus por todos los rincones del planeta, la respuesta de los medios fue consecuente, volcando buena parte de sus recursos en la cobertura de la COVID-19.

Es alrededor de marzo de 2020 cuando surgen los primeros trackers o paneles de datos que ofrecían información diaria a los lectores sobre el avance del virus. Este tipo de piezas

---

<sup>160</sup> Este análisis propio de la automatización en el periodismo visual y los trackers de datos publicados durante la pandemia de la COVID-19 sección también se alimenta de la investigación y conclusiones preliminares de la investigación y el estudio *Effectiveness, clarity and impact of visual journalism in the time of a pandemic* apoyado y financiado por el *Tow Center for Digital Journalism de la Universidad de Columbia*. Para conocer más información sobre el estudio se recomienda la consulta de los anexos: “Anexo I. Formulario para entrevistas a profesionales de la información sobre su trabajo de análisis y la visualización de datos de la COVID-19”, “Anexo II. Artículo preliminar del estudio *Effectiveness, clarity and impact of visual journalism in the time of a pandemic*” y “Anexo III. *Trackers* seleccionados para el estudio de la automatización en la cobertura de la COVID-19”.

visuales se harán un hueco durante meses en las principales portadas de los diarios digitales gracias a la información de servicio público que ofrecen.

Un tracker, por su denominación en inglés, es un panel de datos o pieza periodística rica en datos, normalmente presentados de forma visual, que ofrece información actualizada y cuyas actualizaciones se producen de forma automática. Los trackers pueden tener una utilidad interna para la redacción, por ejemplo, una herramienta que compruebe actualizaciones de datos en una web y muestre un sumario de los mismos al periodista, o externa, por ejemplo, los paneles de datos publicados sobre la COVID-19.

Los trackers de la pandemia son el ejemplo más claro reciente en el que la tecnología y la automatización de procesos han jugado un papel fundamental en el proceso periodístico. La mayoría de las piezas publicadas en los medios digitales que informaban sobre el avance de la pandemia se sirvieron de la automatización para ofrecer información actualizada y rigurosa sin requerir constantemente del periodista para su publicación.

En un ambiente donde los lectores buscaban conocer más detalles sobre un fenómeno nuevo para ellos como es una pandemia, los medios de comunicación reaccionaron ofreciendo información precisa y constante sobre el avance y el impacto del coronavirus en la población.

Al igual que los lectores, los periodistas y profesionales de la información también afrontaron una situación nueva para ellos como era una pandemia. A la vez que aprendían nuevos detalles sobre el virus y la nueva realidad mundial, trabajaban en piezas explicativas sobre la COVID-19. Términos como aplanar la curva o mantener la distancia social se convirtieron en la preocupación diaria tanto de los lectores como de los profesionales de la información.

En esta situación de incertidumbre y de necesidad de reaccionar al ciclo informativo con rapidez y rigurosidad, es donde la automatización de procesos surge como una alternativa válida para ofrecer soluciones a los problemas que afrontaban las redacciones y sus

periodistas. En concreto, los medios de comunicación encontraron una solución en los paneles de datos o trackers para responder a la necesidad de información diaria y constantemente actualizada que demandaban los lectores.

### **Los inicios de la pandemia en la cobertura visual de los medios**

South China Morning Post fue uno de los primeros medios en publicar contenido rico en datos y explicaciones sobre la pandemia. Lo que en un principio comenzó como una cobertura de un centenar de casos derivó en un complejo tracker o panel de datos que actualizaban y actualizan diariamente.

South China Morning Post fue uno de los primeros medios en publicar una pieza visual con datos sobre la pandemia. La pieza, la primera historia visual publicada sobre el coronavirus, se publicó en primer lugar con una serie de gráficos estáticos e ilustraciones.

De acuerdo con los periodistas de South China Morning Post, cuando el coronavirus comenzó a extenderse, decidieron centrar todos sus esfuerzos en la cobertura visual del virus. Debido al conocimiento adquirido durante la cobertura del Síndrome respiratorio agudo grave (SARS) a principios de siglo, intuyeron que la COVID-19 podría convertirse en unos pocos días en una historia global, incluso en una pandemia, como semanas más tarde la Organización Mundial de la Salud (OMS) confirmó.

Mientras el resto del mundo pensaba que el coronavirus sólo era un problema de China, South China Morning Post publicó el primer análisis visual con datos de la COVID-19.

La estrategia que siguieron fue centrarse en un ángulo específico sobre la historia centrándose en la provincia de Wuhan. Publicaron la historia tras un día de trabajo. Con el paso de los días y según la pandemia avanzaba, actualizaron la pieza ampliando la cobertura de los gráficos a China y más tarde a diferentes países como Hong Kong o Tailandia.

Otro de los elementos que añadieron tras unos días fue una breve explicación de cómo el virus puede contagiarse mediante un estornudo o por el aire <sup>161</sup>. Tras unas semanas, ante el abrumador crecimiento de los positivos y contagiados, los periodistas de South China Morning Post tuvieron que pivotar y desarrollar una serie de actualizaciones semiautomáticas en la pieza.

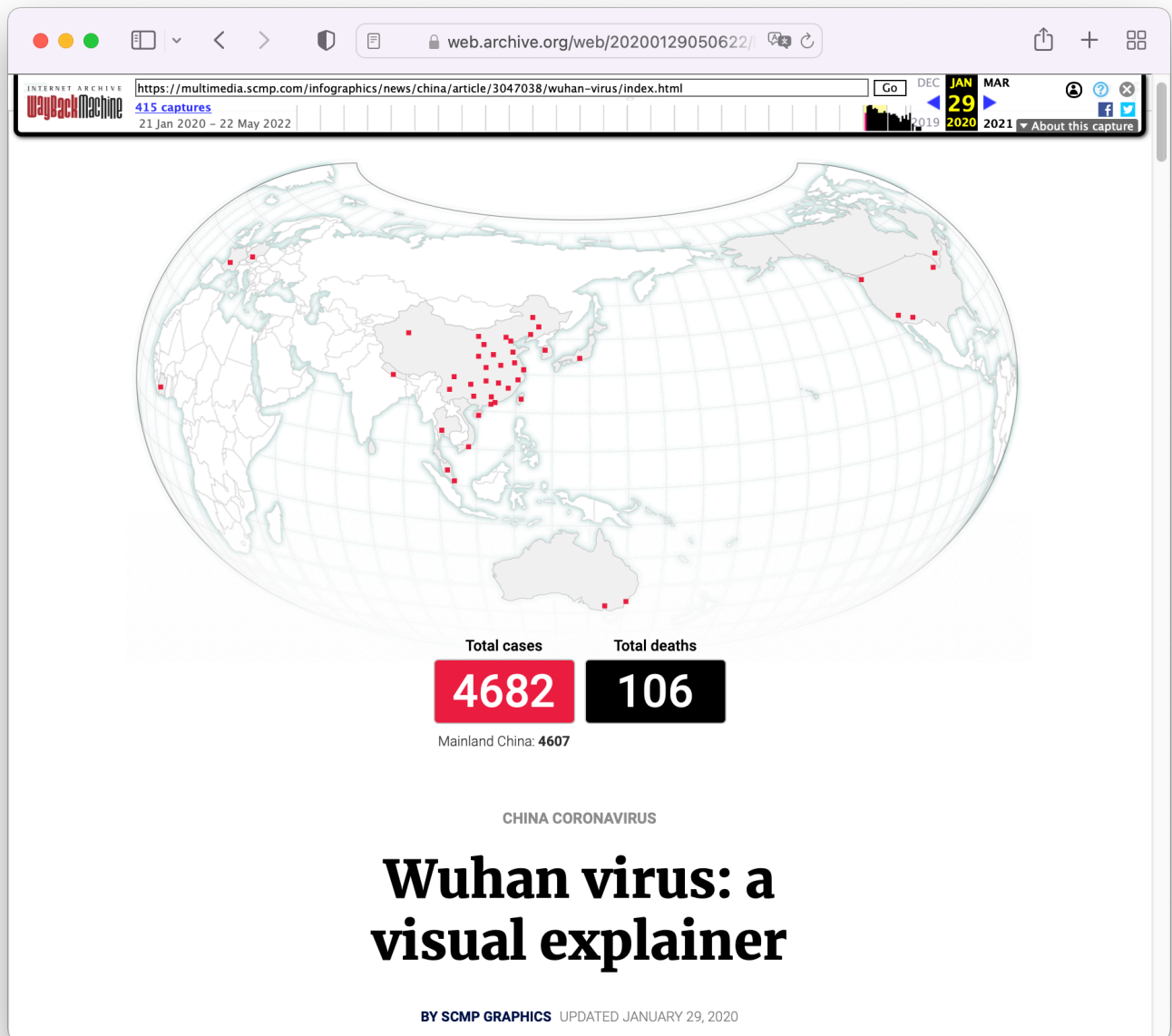
Dicha pieza evolucionará durante las siguientes semanas hasta convertirse en un tracker o panel de datos <sup>162</sup>.

---

<sup>161</sup> Forrest, J. y Wahid, R. (7 de noviembre de 2021). *Interview with the South China Morning Post Infographics team, Nightingale*. Nightingale Journal. Consultado a 15 de diciembre de 2021, disponible en: <https://nightingaledvs.com/on-the-success-of-the-south-china-morning-post-infographics-team/>

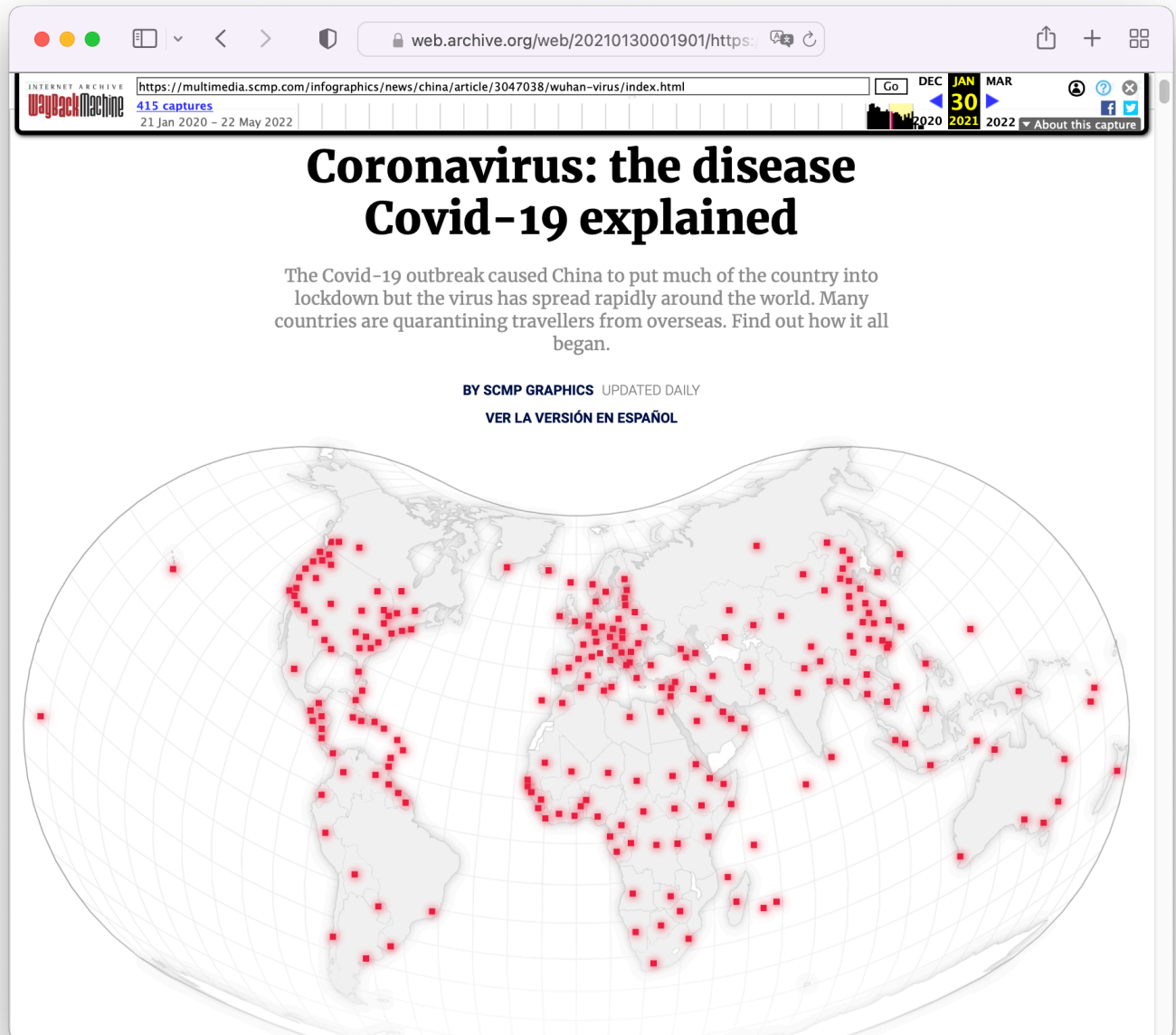
<sup>162</sup> South China Morning Post. (s.f.). *Coronavirus: The disease covid-19 explained*. South China Morning Post. Consultado a 8 de marzo de 2022, disponible en: <https://multimedia.scmp.com/infographics/news/china/article/3047038/wuhan-virus/index.html>

Figura 18. Aspecto del tracker de South China Morning Post de impacto de la COVID-9 el 29 de enero de 2020.



Fuente: Wayback Machina, South China Morning Post.

Figura 19. Aspecto del tracker de South China Morning Post de impacto de la COVID-9 el 30 de enero de 2021.



Fuente: Wayback Machina, South China Morning Post.

A nivel internacional, el resto de medios hizo lo propio y desarrolló una serie de trackers tanto de cobertura global como regional para supervisar y observar el avance del virus e informar a los lectores. Esta apuesta por monitorear el avance de la pandemia con trackers que ofrecen información visual es una iniciativa al alcance de los medios con mayores recursos debido a los conocimientos y profesionales multidisciplinares con los que debe contar una redacción para poder desarrollarla.

Aunque existen un buen número de ejemplos de trackers desarrollados por los medios de comunicación para la cobertura de eventos políticos <sup>163</sup>, económicos <sup>164</sup> o de verificación de datos <sup>165</sup>, entre otros, la cobertura de la pandemia se convirtió en el escenario perfecto para este tipo de piezas periodísticas en las que la automatización juega un papel clave.

Para Danzon-Chambaud, el volumen de datos liberado durante la pandemia y el interés de los lectores por estar informados de la propagación y efectos del virus en su entorno convierten a la cobertura de la COVID-19 en la historia ideal para aplicar los fundamentos de la automatización de procesos <sup>166</sup>:

“En cierto modo, la COVID-19 puede ser la perfecta historia automatizar. Según el virus se expandió por el planeta a comienzos de 2020, los gobiernos y las

---

<sup>163</sup> Stiles, M. (1 de marzo de 2020). *Track donations in the L.A. county supervisor's race for the 2nd district*. Los Angeles Times. Consultado a 12 de enero de 2022, disponible en: <https://www.latimes.com/projects/county-supervisor-2nd-district-campaign-donations/>

<sup>164</sup> Willis, D. y Syed, M. (7 de Julio de 2020). *Tracking PPP: Search every company approved for Federal Loans*. ProPublica. Consultado a 12 de enero de 2022, disponible en: <https://projects.propublica.org/coronavirus/bailouts/>

<sup>165</sup> The Washington Post. (20 de enero de 2021). *Tracking all of president Trump's false or misleading claims*. The Washington Post. Consultado a 12 de enero de 2022, disponible en: <https://www.washingtonpost.com/graphics/politics/trump-claims-database/>

<sup>166</sup> Danzon-Chambaud, S. (17 de septiembre de 2021). *How news organizations used Automated News to cover COVID-19*. Poynter. Consultado a 13 de enero de 2022, disponible en: <https://www.poynter.org/tech-tools/2021/how-news-organizations-used-automated-news-to-cover-covid-19/>

autoridades sanitarias liberaron una cantidad considerable de datos para el público, como son el número de fallecidos, el número de pacientes en unidades de cuidados intensivos y la incidencia acumulada semanal”.

Para analizar las características comunes de un tracker y el impacto acelerador que la pandemia ha tenido en la automatización de procesos en el periodismo visual, esta tesis examina una docena de trackers que han sido publicados durante la pandemia para informar al lector de visual del avance del virus <sup>167</sup>.

Con el fin de realizar una comparación justa se han seleccionado una docena de trackers que muestran el avance del virus a nivel mundial, desarrollados por medios grandes que cuentan con perfiles multidisciplinares o equipos de Datos, Gráficos e Ingeniería en sus redacciones. Para la selección y estudio de esos trackers se han seguido una serie de criterios con el fin de comparar de forma objetiva los diferentes elementos de cada uno de ellos.

- Criterio de automatización: Algunos o todos los procesos de procesamiento, análisis y visualización de datos del tracker han sido automatizados por los profesionales de la información.
- Criterio de contenidos: Los trackers seleccionados analizan al menos dos variables del virus (casos positivos y fallecidos) y cuentan con al menos tres tipos de visualizaciones, entre gráficos y mapas, en la pieza periodística.
- Criterio de diversidad geográfico: Dentro de lo posible, se ha buscado variedad geográfica a la hora de seleccionar los medios con el fin de comparar los trackers desarrollados por medios de comunicación de diferentes continentes.
- Criterio de relevancia y popularidad: Se han seleccionado algunos de los trackers más populares para los lectores y los expertos de la visualización de datos.

---

<sup>167</sup> Ver “Anexo III. *Trackers* seleccionados para el estudio de la automatización en la cobertura de la COVID-19” para explorar la base de datos elaborada para el estudio y análisis de las diferentes características de cada uno de los *trackers* estudiados en esta tesis.

- Criterio de tamaño: se han seleccionado medios de similares características y tamaño. Todos ellos son medios grandes con recursos.

Una vez seleccionados los criterios, se ha elaborado una base de datos sencilla con los trackers y se han analizado los diferentes elementos de cada uno de ellos. En el “Anexo III. Trackers seleccionados para el estudio de la automatización en la cobertura de la COVID-19” se incluye el detalle y las conclusiones de cada uno de los elementos analizados a marzo de 2022.

Las principales conclusiones extraídas de la intersección entre periodismo robot y periodismo visual, así como los efectos que la pandemia ha tenido en estas piezas visuales siguen a continuación.

### **Desafíos con los datos**

Una de las limitaciones en el uso de la automatización es la falta de supervisión humana en las actualizaciones periódicas de los datos o los reportajes. Es uno de los problemas que han tenido que afrontar cientos de profesionales de datos a la hora de programar, diseñar y visualizar las diferentes piezas que incluían un sumario con los datos de la pandemia.

La automatización de procesos choca en ocasiones con algunos de los principales objetivos del periodismo como son la precisión y el rigor. Al programar un análisis de datos, una visualización o una pieza periodística, el profesional de la información debe tener en cuenta los diferentes problemas que el robot que ha escrito puede encontrarse en el futuro cuando esté ejecutando los procesos por su cuenta sin la supervisión del periodista.

En la mayoría de los casos, la opción más sencilla a la hora de programar estos robots es utilizar cláusulas condicionales que tengan en cuenta los diferentes escenarios inesperados que puedan surgir. Por ejemplo, uno de los casos problemáticos en las primeras semanas

desde que se declaró el estado de pandemia mundial fueron los múltiples formatos y estructuras en los que los datos estaban disponibles, así como los errores que de vez en cuando se escondían en los mismos.

El departamento *The Center for Systems Science and Engineering (CSSE)* de *John Hopkins University* en Estados Unidos fue una de las primeras instituciones de prestigio en crear una base de datos estructurada y accesible a nivel mundial de casos, fallecidos y recuperados de la COVID-19. Por ello, se convirtió durante las primeras semanas en la fuente de datos por excelencia a la que acudir para informar sobre la pandemia.

En los inicios de la pandemia, el trabajo hercúleo del equipo de epidemiología de *John Hopkins University* consistió en crear una base de datos con las cifras de la COVID-19 por países. Su labor diaria consistía en buscar, chequear e introducir el número de casos, fallecidos y recuperados de la COVID-19 por país en una base de datos. Posteriormente, liberaron en la plataforma GitHub <sup>168</sup> dicha base de datos que comparaba el avance del virus por países.

Las manos eran limitadas en el equipo de la universidad americana y la cantidad de países con un número preocupante de casos comenzó a ser abrumador a principios de abril. Por consiguiente, la probabilidad de cometer un error humano al introducir un número en la base de datos era cada vez más tangible <sup>169</sup>.

Además, otro de los problemas con los que contaban los científicos de *John Hopkins University* era la disparidad de criterios y metodologías entre países a la hora de liberar y publicar los datos.

---

<sup>168</sup> CSSEGISandData. (s.f.). *CSSEGISANDDATA/covid-19: Novel coronavirus (COVID-19) cases, provided by JHU CSSE*. GitHub. Consultado a 21 de enero de 2022, disponible en: <https://github.com/CSSEGISandData/COVID-19>

<sup>169</sup> Swenson, Kyle. (29 de abril de 2020). *Millions track the pandemic on Johns Hopkins's dashboard. Those who built it say some miss the real story*, *The Washington Post*. Consultado a 17 de enero de 2022, disponible en: [https://www.washingtonpost.com/local/johns-hopkins-tracker/2020/06/29/daea7eea-a03f-11ea-9590-1858a893bd59\\_story.html](https://www.washingtonpost.com/local/johns-hopkins-tracker/2020/06/29/daea7eea-a03f-11ea-9590-1858a893bd59_story.html)

De esta forma, con el paso de las semanas, el equipo de John Hopkins University se vio obligado a introducir una serie de cambios y correcciones en la metodología de sus datos, así como en la estructura de los mismos para facilitar la recogida y publicación de los mismos <sup>170</sup>.

Estos cambios en los datos son importantes para entender las soluciones que los profesionales de la información han aplicado para automatizar las piezas sobre el avance de la pandemia que tan útiles han sido para el lector en los primeros meses de la propagación del virus.

En el análisis de los diferentes trackers publicados por medios nacionales e internacionales durante los dos últimos dos años pueden verse algunas estrategias para reducir los problemas y desafíos que la automatización todavía conlleva en el mundo del periodismo. Estas disparidades en la metodología y publicación de los datos del coronavirus por parte de organizaciones, administraciones y autoridades sanitarias ha sido un quebradero de cabeza para los periodistas y ha generado varios desafíos a la hora de programar los diferentes trackers.

Al comienzo de la pandemia, la mayoría de los medios de comunicación analizados utilizaron los datos publicados por Johns Hopkins University por ser los únicos disponibles. Con el paso de los meses, reemplazaron dichos datos por los datos proporcionados por otras fuentes como Our World in Data, una iniciativa de la Universidad de Oxford para publicar

---

<sup>170</sup> CSSEGISandData. (s.f.). *Covid-19/archived\_data/archived\_daily\_case\_updates at master · CSSEGISANDDATA/covid-19*. GitHub. Consultado a 21 de enero de 2022, disponible en: [https://github.com/CSSEGISandData/COVID-19/tree/master/archived\\_data/archived\\_daily\\_case\\_updates](https://github.com/CSSEGISandData/COVID-19/tree/master/archived_data/archived_daily_case_updates)

datos estructurados de la COVID-19, o incluso por los datos recopilados por sus propios reporteros como detallan en las fuentes de datos al pie de estas piezas visuales <sup>171</sup>.

## **Flexibilidad en el diseño y en el uso de variables**

Uno de los momentos críticos a la hora de crear un tracker es el diseño de sus características y la toma de decisiones al comienzo de su creación teniendo en cuenta posibles cambios o ediciones futuras.

El diseño computacional previo de los elementos, características y funcionalidades de un tracker es fundamental del mismo modo que el diseño de las características de un algoritmo es clave para su correcto funcionamiento como se ha explicado en el capítulo anterior. Bien es cierto que el desarrollo de trackers o paneles de datos permite cierta flexibilidad una vez publicado el proyecto debido al carácter cambiante de los datos y de las informaciones con incertidumbre.

A comienzos de 2020, los trackers analizados en esta tesis informaban sobre el número de tests realizados en la población, del número de casos positivos y de las personas fallecidas y recuperadas. Con el paso de las semanas, los medios de comunicación pivotaron ligeramente y se centraron únicamente en el número de casos y fallecidos debido a que representaban de forma más fehaciente el avance de la pandemia. Un año después del inicio de la pandemia, los medios de comunicación incorporaron el número de personas vacunadas a sus trackers o crearon trackers centrados exclusivamente en informar sobre el ratio de vacunación en los diferentes países.

El razonamiento detrás de este ajuste en la información y variables de la COVID-19 procesada y proporcionada por los medios de comunicación busca representar de forma más exacta el avance y los efectos de la COVID-19 en la población.

---

<sup>171</sup> Blanco, A. y Sauras, J. (1 de junio de 2021). *Drawing the curve: Data visualization and COVID-19 A pandemic year in graphics*. Tow Center for Digital Journalism at Columbia's Graduate School of Journalism. Consultado a 2 de junio de 2021, disponible en: <https://mailchi.mp/95bdcee6e0d7/weekly-updates-on-covid-19s-impact-on-journalism-2083305?e=e8e508e35f>

## **Incertidumbre**

La limitada disponibilidad de datos al comienzo de la pandemia, los errores en los mismos y las constantes correcciones y ediciones que las fuentes de información realizaron, dieron paso a la cuestión de cómo representar esa incertidumbre en las visualizaciones de los trackers.

Como Hullman ha estudiado, la presencia de la incertidumbre en las piezas visuales de datos publicadas en los medios de comunicación suele ser una excepción más que una regla <sup>172</sup>.

En este sentido, el reducido uso de técnicas para mostrar la incertidumbre y la variabilidad de los datos no es una excepción en los trackers analizados. La mayoría de los medios de comunicación apostaron por acompañar estas piezas visuales ricas en datos con extensas metodologías o notas al pie de las visualizaciones donde advierten de los problemas que pueden presentar los datos.

Además, todos los trackers analizados incluyen una media semanal de casos o fallecidos en forma de gráfico de líneas con el objetivo de representar de forma más certera la tendencia, al alta o la baja, en la evolución de la pandemia.

*The Washington Post*, por su parte, apostó por evitar la publicación de números muy detallados y redondeó a la baja en la cifra de los millares el número de casos y fallecidos con el objetivo de expresar la incertidumbre e imprecisión de los datos disponibles. Tanto en el titular como en la pieza, todos los números están redondeados y el uso de expresiones como “más de” o “al menos” es constante.

---

<sup>172</sup> Hullman, J. (2020). *Why Authors Don't Visualize Uncertainty*, IEEE Transactions on Visualization and Computer Graphics, 26, 130-139, disponible en: <https://ieeexplore.ieee.org/document/8805422>

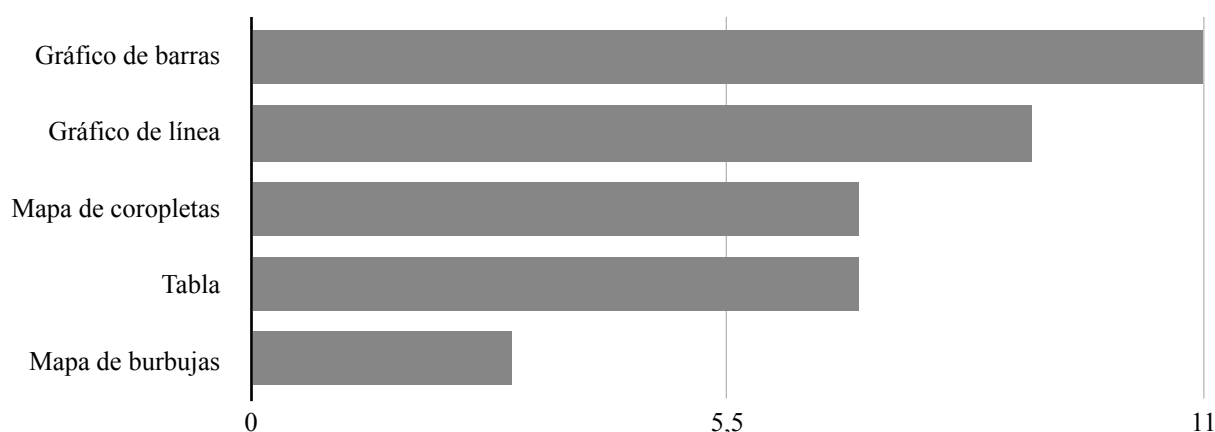
## El uso de los gráficos y mapas

Otro de los elementos clave a la hora de automatizar los procesos que estos trackers ejecutan para ofrecer información al lector fue el tipo de gráfico elegido para transmitir la información de forma sencilla y clara. Para categorizar cada gráfico en función de su aspecto se ha utilizado como referencia las taxonomías elaboradas por Hildén y Koponen <sup>173</sup> y por Wille <sup>174</sup>.

Los trackers estudiados en esta tesis muestran la apuesta de los profesionales de la información por visualizaciones sencillas que pueda comprender cualquier lector. Los gráficos de barras y de líneas son las principales apuestas en el diseño de estas piezas visuales para mostrar la evolución de casos y fallecidos.

Existe una mayor variedad en el uso de los mapas y las visualizaciones espaciales. Pese a ello, los mapas de coropletas y de burbujas son las opciones más repetidas en los diferentes trackers analizados.

Figura 20. Tipos de gráficos y mapas utilizados en los diferentes trackers estudiados en esta disertación.



<sup>173</sup> Hildén, J., Koponen, J. (2019). *Data Visualization Handbook*. Finland: Aalto University, pp. 137-230.

<sup>174</sup> Wilke, C. (2019). *Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures*. Taiwan: O'Reilly Media, pp. 37-44.

Fuente: Elaboración propia a partir del análisis realizado de los trackers del coronavirus.

Danzon-Chambaud destaca los beneficios que el uso de los procesos de automatización ha tenido en los medios de comunicación tras estudiar nueve iniciativas periodísticas, siete europeas y dos norteamericanas. En primer lugar, destaca el uso de la automatización para cuantificar la pandemia y mostrar contenido actualizado y preciso al lector sobre el avance de la COVID-19.

En segundo lugar, el autor destaca el uso interno de la automatización en los medios de comunicación para analizar datos de la COVID-19 e identificar y desarrollar nuevas historias <sup>175</sup>.

Además, Danzon-Chambaud analiza el impacto que puede tener “este matrimonio entre humanos y máquinas” que puede derivar en la producción veloz de contenidos de calidad gracias al soporte de la automatización de procesos, mejorando el posicionamiento de los medios en internet y su alcance a nuevos lectores y audiencias <sup>176</sup>.

#### **5.4. Humanos y máquinas para cubrir la pandemia, el caso de *The COVID Tracking Project***

Con el paso de los meses, una de las iniciativas periodísticas que más impacto tuvo en Estados Unidos a la hora de automatizar y procesar datos sobre la COVID-19 fue *The Covid Tracking Project*. Esta iniciativa es el ejemplo por excelencia del efecto que ha tenido la pandemia a la hora de acelerar el uso de procesos tecnológicos en el sector del periodismo.

---

<sup>175</sup> Danzon-Chambaud, S. (6 de agosto de 2021). *Covering covid-19 with Automated News*. Columbia Journalism Review. Consultado a 14 de enero de 2022, disponible en: [https://www.cjr.org/tow\\_center\\_reports/covering-covid-automated-news.php](https://www.cjr.org/tow_center_reports/covering-covid-automated-news.php)

<sup>176</sup> Ibid.

A comienzos de marzo de 2020, Robinson Meyer y Alexis Madrigal, dos periodistas del medio estadounidense *The Atlantic*, comenzaron a recopilar datos del coronavirus para una de sus historias sobre la escasez y la falta de coordinación para realizar tests a la población americana. Trabajando en este proyecto, ambos reporteros identificaron una necesidad crítica en ese momento como era la disponibilidad de datos detallados y disponibles públicamente para poder analizar y contar qué estaba pasando. Este hecho les llevó a fundar *The COVID Tracking Project*. A la iniciativa de recogida de datos sobre el coronavirus se uniría más tarde el científico de datos y fundador de la startup en el sector de la salud Related Sciences, Jeff Hammerbacher.

Es en este momento cuando realizaron una llamada de ayuda a todos aquellos voluntarios que quisieran unirse a contribuir al esfuerzo de recopilar datos. Al mismo tiempo llega Erin Kissane que hasta entonces había dirigido el departamento de contenidos de la fundación Knight-Mozilla OpenNews, para dirigir el proyecto.

La iniciativa surgió de manera informal a principios de marzo con un pequeño equipo que trabajaba con una hoja de cálculo donde recogían información y datos del impacto de la COVID-19 por estado. Ese equipo minúsculo creció hasta albergar cientos de voluntarios que trabajaron durante más de un año recopilando datos de forma manual a diario. El perfil de estos voluntarios va desde recolectores de datos, a científicos, desarrolladores, diseñadores, periodistas y editores. El proyecto ejemplifica una redacción moderna donde es cada vez más común encontrar perfiles multidisciplinares y diversos.

Los editores de *The COVID Tracking Project* explican que utilizaron la automatización de procesos para ayudar y apoyar el trabajo manual que los cientos de voluntarios que han formado parte del proyecto realizaron durante meses. En ningún momento trataron de reemplazar a estos voluntarios, base del proyecto, mediante la

automatización ya que querían asegurarse de que tenían el control de los datos y, por ende, las conclusiones que extraían de los mismos <sup>177</sup>.

Otros aspectos clave que muchos profesionales de la información tuvieron que afrontar en la cobertura del coronavirus fue la dificultad para acceder a datos fiables y la disparidad de criterios que siguieron las diferentes administraciones para publicar estas bases de datos.

Tras las primeras semanas de propagación del virus, algunas autoridades sanitarias de Estados Unidos comenzaron a publicar paneles de datos sobre la COVID-19 en sus páginas web, mientras otras divulgaron las cifras más relevantes en ruedas de prensa. En algunos casos, incluso, el apagón informativo por parte de la administración fue total respecto al impacto del coronavirus en los diferentes estratos de población ya que no estaban organizadas o carecían de los recursos necesarios para recopilar y publicar datos.

Al igual que en el resto del mundo, los reporteros y colaboradores de *The COVID Tracking Project* tuvieron que superar estos desafíos. Con la experiencia adquirida, pronto se dieron cuenta de que para poder ofrecer datos rigurosos tenían que estar en contacto directo con las fuentes de los datos o administraciones en posesión de los registros para entender qué estaba ocurriendo <sup>178 179</sup>:

---

<sup>177</sup> Gilmour, J. (28 de mayo de 2021). *Analysis & Updates: 20,000 hours of data entry: Why we didn't Automate Our Data Collection*. The COVID Tracking Project. Consultado a 16 de enero de 2022, disponible en: <https://covidtracking.com/analysis-updates/why-we-didnt-automate-our-data-collection>

<sup>178</sup> Allsop, J. (17 de marzo de 2021) *The COVID Tracking Project is (nearly) gone. Can we see clearly now*, Columbia Journalism Review (CJR). Consultado a 21 de enero de 2022, disponible en: [https://www.cjr.org/the\\_media\\_today/the-covid-tracking-project-is-nearly-gone-can-we-see-clearly-now.php](https://www.cjr.org/the_media_today/the-covid-tracking-project-is-nearly-gone-can-we-see-clearly-now.php)

<sup>179</sup> CNN (16 de marzo de 2021). *America wasn't ready for Covid-19. These newsrooms helped guide the way*, CNN. Consultado a 21 de enero de 2022, disponible en: <https://www.cnn.com/2021/03/16/media/the-atlantic-stat-pandemic-journalism/index.html>

“Al comienzo, pensamos que podríamos extraer datos de los paneles de datos públicos y que nuestro trabajo se resumiría en eso. Como se vio después, lo que realmente necesitábamos era una investigación profunda de los datos disponibles”.

Durante la pandemia, la publicación y acceso a los datos sobre la COVID-19 ha sido uno de los mayores problemas para los periodistas que han trabajado en su cobertura. Uno de los contratiempos más habituales han sido los constantes cambios en la forma y el tipo de dato divulgado. Además, los constantes cambios en las plataformas y lugares de publicación de esas bases de datos han sido un quebradero de cabeza constante para los reporteros. En el caso de Estados Unidos, ante la falta de una normativa federal para la publicación de los datos, los estados tenían la última decisión en relación a cómo liberar los datos para disposición del público.

En cualquier caso, a pesar de las precauciones a la hora de recopilar los datos y la necesidad de realizar un recogida y comprobación manual de los mismos, el equipo de *The COVID Tracking Project* automatizó un buen número de procesos que facilitaron la labor de las docenas de reporteros que trabajaron en el proyecto.

Entre las herramientas de automatización de procesos creadas por *The COVID Tracking Project* destacan las siguientes.

Por un lado, los profesionales que trabajaron en el proyecto crearon una serie de robots que capturaba la pantalla de los paneles de datos online de la COVID-19 DE los diferentes estados de Estados Unidos. Para ello, configuraron docenas de robots que ejecutaban capturas de pantalla <sup>180</sup> y crearon un archivo de imágenes con datos diarios que les sirvió para verificar los datos recogidos de forma sencilla y accesible para todos los colaboradores. *The COVID Tracking Project* liberó de manera pública dichas capturas de

---

<sup>180</sup> The COVID Tracking Project. (s.f.). *State screenshots*. The COVID Tracking Project. Consultado a 4 de febrero de 2022, disponible en: <https://screenshots.covidtracking.com/>

pantalla de los paneles del avance de la COVID-19 para que cualquier lector o usuario pudiera utilizarlos.

Esta herramienta de automatización de procesos es similar en su concepción a la plataforma Wayback Machine <sup>181</sup> que pretende crear un archivo de internet mediante capturas diarias de cientos de miles de páginas web.

Aunque los datos que ofrece *The COVID Tracking Project* eran recogidos y chequeados manualmente por sus colaboradores, también crearon un robot que recopila y descarga de forma automática los datos de cada uno de los portales de estatales con datos de la COVID-19.

Esta herramienta fue clave para poder verificar los datos recogidos por los voluntarios y, de esta forma, evitar errores. En total, el trabajo de los voluntarios supuso más de 20,754 horas de trabajo anuales y cientos de valores y datos individuales introducidos en la base de datos a diario <sup>182</sup>. Además, en algunos casos, los voluntarios simplemente tenían que verificar y validar los datos recogidos por el robot, en lugar de introducir los datos en una hoja de cálculo.

La labor de otro de los robots que creó el equipo de *The COVID Tracking Project* fue la de procesar los datos que previamente habían recopilado y analizarlos, calculando casos acumulativos y diarios en función de las variables disponibles. Esta herramienta les alertaba de posibles errores en caso de que las cifras fuesen muy dispares a lo esperado.

También escribieron y crearon robots que les ayudaron a recopilar datos publicados a través de herramientas propietarias como Power BI dashboards o ArcGIS dashboard. Estas

---

<sup>181</sup> Wayback Machine. (s.f.) Wayback Machine. Consultado a 4 de febrero de 2022, disponible en: <https://web.archive.org/>

<sup>182</sup> French, A., & Camberg, N. (22 de abril de 2021). *Inside the Covid Tracking Project's volunteer organization*. The COVID Tracking Project. Consultado a 23 de febrero de 2022, disponible en: <https://covidtracking.com/analysis-updates/inside-the-covid-tracking-projects-volunteer-organization>

herramientas, muy utilizadas por los gobiernos y administraciones en la mayoría de los países, suponen un quebradero de cabeza adicional para el profesional de la información ya que no permiten la descarga directa de los datos. Recopilar los datos de estas herramientas supone muchas veces una labor tediosa y manual.

Por último, pese a que la automatización no era fiable para hacerse cargo de todo el proceso, sí fue de mucha ayuda para capturar y registrar los cambios constantes en la variable temporal de los datos.

Las administraciones sanitarias en América y en el resto del mundo <sup>183</sup>, cambiaron la periodicidad de las series de datos múltiples veces en los dos últimos años. Por otro lado, el retraso en la publicación de ciertos datos ha sido constante especialmente durante los períodos vacacionales y los fines de semana <sup>184 185</sup>.

De esta forma, días o semanas más tarde de la publicación de un conjunto de datos, las administraciones revisaban y ajustaban los datos publicados anteriormente.

Mediante un robot que era capaz de extraer datos e información en una fecha específica, *The COVID Tracking Project* fue capaz de corregir cualquier error, cambio o discrepancia en los datos días más tarde de que estos se hubieran publicado y validado <sup>186</sup>.

---

<sup>183</sup> Neutral. (9 de marzo de 2022). *Sanidad dejará de publicar los datos sobre evolución del covid a diario*. Newtral. Consultado a 12 de marzo de 2022, disponible en: <https://www.newtral.es/sanidad-datos-covid-informe/20220305/>

<sup>184</sup> Harvard T.H. Chan School Of Public Health. (14 de julio de 2020). *Data animation shows time lag between COVID-19 cases and deaths*. Harvard T.H. Chan School of Public Health. Consultado a 10 de febrero de 2022, disponible en: <https://www.hsph.harvard.edu/news/hsph-in-the-news/data-animation-shows-time-lag-between-covid-19-cases-and-deaths/>

<sup>185</sup> The New York Times. (22 de noviembre de 2021) *The Trouble With the Case Curve During the Holidays*. The New York Times. Consultado a 10 de febrero de 2022, disponible en: <https://www.nytimes.com/interactive/2021/11/22/us/covid-data-holiday-averages.html>

<sup>186</sup> Michel, T. Y Rebma. (13 de mayo de 2021). *Analysis & Updates: Dating data: How we used multiple dating schemes to provide the most complete picture of the pandemic*. The COVID Tracking Project. Consultado a 10 de febrero de 2022, disponible en: <https://covidtracking.com/analysis-updates/how-we-used-multiple-dating-schemes-to-provide-the-most-complete-picture-of-the-pandemic>

Con el avance de la pandemia y según la publicación de los datos y las metodologías se afianzaron y comenzaron a ser más estables, *The COVID Tracking Project* prosiguió automatizando más procesos ya que el margen de cometer un error disminuyó. Bien es cierto que, como señalan, nunca automatizaron el proceso de recogida, validación y publicación de los datos. En especial, los periodistas de *The COVID Tracking Project* fueron muy cuidadosos en la validación y publicación de los datos y las conclusiones — siempre supervisadas por un periodista, voluntario o profesional de la información <sup>187</sup>.

### **5.5. Automatización y bases de datos propias, la estrategia de *The New York Times* en la cobertura de la pandemia de la COVID-19**

*The New York Times* es uno de los medios que más ha aplicado la automatización de procesos a la hora de visualizar y presentar los datos de la COVID-19 como puede verse en los numerosos trackers o paneles de información que ofrecen a sus lectores <sup>188</sup>.

Al igual que *The Covid Tracking Project*, los periodistas de *The New York Times* crearon una estructura y metodología para recopilar los datos sobre la pandemia que gobiernos y autoridades sanitarias han publicado desde principios de 2020. Por su parte, la cobertura de *The New York Times* destaca por la variedad de sus contenidos y la multitud de datos relacionados con la COVID-19 que han recopilado.

Las fuentes de datos que potencian las piezas y paneles de datos que ofrece *The New York Times* destacan por su variedad. Los datos sobre el número de casos y fallecidos en Estados Unidos provienen de los informes de las autoridades locales y estatales estadounidenses. Los datos sobre hospitalizaciones y número de test provienen de *The Covid Tracking Project*. Por último, los datos internacionales que utilizan en sus piezas son los

---

<sup>187</sup> The COVID Tracking Project. (s.f.). *About Us*. Consultado a 10 de febrero de 2022, disponible en: <https://covidtracking.com/about>

<sup>188</sup> The New York Times. (24 de noviembre de 2020). *Track coronavirus cases in places important to you*. The New York Times. Consultado a 11 de febrero de 2022, disponible en: <https://www.nytimes.com/interactive/2021/us/covid-cases-deaths-tracker.html>

datos recopilados y liberados por *John Hopkins University*. Además, los periodistas de *The New York Times* han creado sus propias bases de datos y trackers sobre el impacto que ha tenido la COVID-19 en las universidades <sup>189</sup> y las residencias de ancianos <sup>190</sup>.

Parte del proceso de análisis, visualización y publicación de los paneles de datos que *The New York Times* ofrece ha sido automatizado por sus periodistas con el fin de ofrecer información actualizada cada hora.

En cualquier caso, un análisis de los elementos presentes en las piezas semiautomatizadas publicadas por el diario estadounidense revela su aproximación a la automatización, así como los beneficios de aplicar estas técnicas al procesamiento, visualización y publicación de datos en periodismo.

En primer lugar, cabe destacar que no todo el proceso de recolección, análisis y publicación de datos es completamente automático. Un primer indicio de este hecho es el número de periodistas en plantilla que firman estas informaciones visuales.

En los trackers publicados, *The New York Times* distingue entre los diferentes roles que han jugado sus periodistas al trabajar en la información:

- 38 profesionales, la mayoría de ellos periodistas de datos, reporteros gráficos e infografistas, firman la pieza principal de seguimiento de la pandemia por su trabajo en el análisis de datos, la visualización y la presentación de las conclusiones y actualizaciones diarias.

---

<sup>189</sup> The New York Times. (25 de febrero de 2021). *Tracking coronavirus cases at U.S. colleges and Universities*. The New York Times. Consultado a 11 de febrero de 2022, disponible en: <https://www.nytimes.com/interactive/2021/us/college-covid-tracker.html>

<sup>190</sup> The New York Times. (27 de junio de 2020). *Nearly one-third of U.S. coronavirus deaths are linked to nursing homes*. The New York Times. Consultado a 11 de febrero de 2022, disponible en: <https://www.nytimes.com/interactive/2020/us/coronavirus-nursing-homes.html>

- 87 reporteros y editores firman la pieza por su contribución investigando, verificando y acudiendo a las fuentes para aclarar cuestiones variadas con los datos.
- 33 profesionales firman la información por su trabajo en la adquisición, recolección y registro de los datos en la base de datos de *The New York Times*.

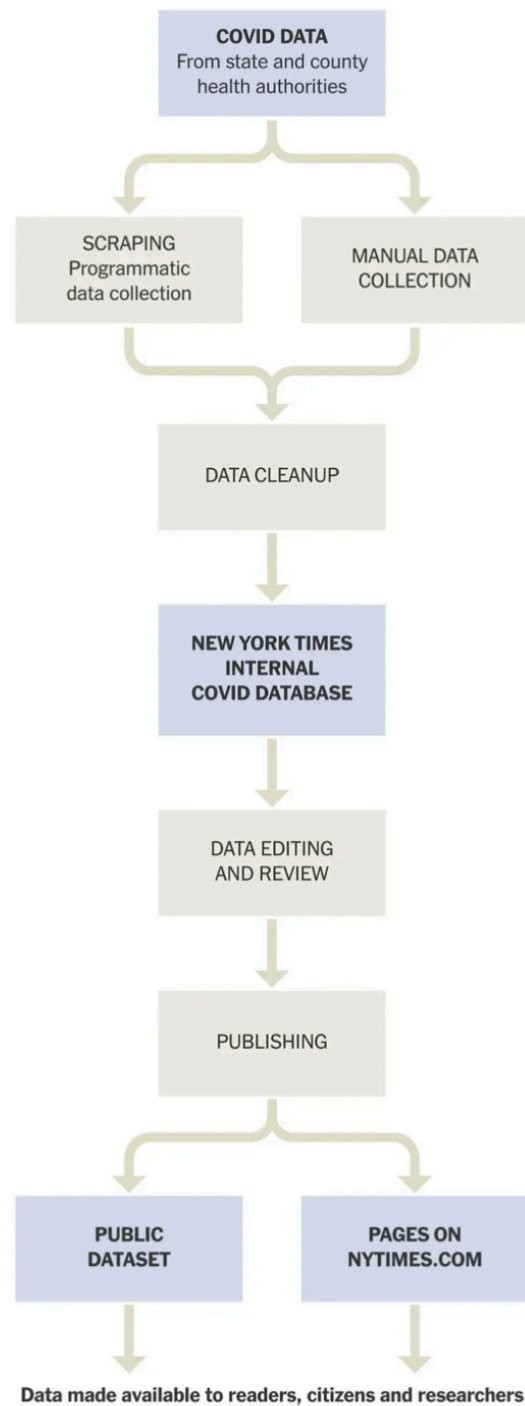
Por ende, se entiende que el paso más delicado para la publicación de estos paneles de datos y piezas informativas como es la recogida, introducción y análisis de los datos está realizado o supervisado por los periodistas de *The New York Times*.

De hecho, de acuerdo con el esquema publicado por el periódico estadounidense, sus periodistas combinan el trabajo manual de recolección de datos con una serie de scripts o robots que descargan los datos de forma automática <sup>191</sup>.

---

<sup>191</sup> Fehr, T. y Williams, J. (24 de junio de 2021). *10 million data requests: How our Covid team tracked the pandemic*. The New York Times. Consultado a 13 de febrero de 2022, disponible en: <https://www.nytimes.com/2021/06/24/insider/covid-tracking-data.html>

Figura 21. Esquema de tareas manuales y automáticas realizadas por los periodistas del diario *The New York Times* para el procesamiento, edición y publicación de datos de la COVID-19.



Fuente: Guilbert Gates/*The New York Times*.

Otra práctica loable por parte de *The New York Times* es la publicación de los datos limpios y estructurados en un repositorio de GitHub para el uso del público, una vez han sido validados por sus tests automáticos y por los periodistas asignados a cubrir el coronavirus. El fomento de la transparencia en este sentido y la posibilidad de que cualquier persona pueda acceder a los datos no sólo ofrece una mayor credibilidad al medio, sino que añade una nueva capa de verificación de los datos ya que cualquier usuario puede notificar posibles errores u ofrecer recomendaciones.

Un dato curioso sobre las numerosas actualizaciones diarias de los datos que realiza The New York Times en GitHub es que un bot realiza y ejecuta dichas tareas de subida de datos. Otro ejemplo de cómo un robot sencillo de programar puede liberar al periodista de una labor que tendría que recordar varias veces al día y que le quitaría un tiempo valioso para invertir en otro tipo de labores como análisis más elaborados de los datos.

En este caso, el código del bot no es abierto, pero puede verse como en el último año el bot <sup>192</sup> ha realizado más de 1.000 actualizaciones por sí mismo <sup>193</sup>. Un hecho que demuestra que incluso las acciones más simples, como es la publicación de una serie de datos, pueden suponer decenas de horas de trabajo a lo largo de un año. Efectivamente, como demuestra este pequeño ejemplo, es en las acciones repetitivas donde la automatización de procesos ha ofrecido hasta ahora un mayor número de soluciones y ahorro de tiempo al periodista.

---

<sup>192</sup> The New York Times. (s.f.). *NYT-COVID-19-bot - overview*. GitHub. Consultado a 15 de febrero de 2022, disponible en: <https://github.com/nyt-covid-19-bot>

<sup>193</sup> The New York Times. (s.f.). *Nytimes/COVID-19-data: An ongoing repository of data on coronavirus cases and deaths in the U.S.* GitHub. Consultado a 15 de febrero de 2022, disponible en: <https://github.com/nytimes/covid-19-data>

## **5.6. Simulaciones y juegos, el papel de la automatización**

Además de la publicación de trackers para informar al lector, la automatización y los algoritmos han permitido desarrollar una serie de géneros de noticia novedosos en el periodismo visual. La simulación de escenarios para explicar un hecho complejo y los newsgames o juegos con vocación informativa destacan entre estas piezas informativas.

Ambos géneros comparten una serie de características en su concepción y confección como son el uso de los datos para construir la pieza, la alta presencia del lenguaje visual y la interactividad de los elementos en la publicación periodística final. Las técnicas de automatización de procesos han hecho posible desarrollar este tipo de piezas que, una vez publicadas, viven por sí solas en la web, mostrando diferentes escenarios en cada lectura del usuario.

La creación de este tipo de piezas supone el conocimiento de diversas técnicas multidisciplinares. El profesional de la información recurre a sus conocimientos de la narrativa para crear la historia, hace uso de los principios de diseño para conceptualizar la pieza y ejecuta el producto final escribiendo lenguaje informático que posibilita la interactividad de la pieza.

### **Simulaciones**

Las simulaciones, artículos interactivos con un alto componente visual, muestran al lector diferentes escenarios basados en datos para explicar un problema o hecho complejo. Además, un buen número de estas piezas permiten que el lector interactúe con los datos y visualice los posibles escenarios que pueden darse en función de la aplicación de diferentes variables o hechos.

Estas piezas suelen contener un componente de abstracción como puede ser un mundo o un país imaginario con el objetivo de simplificar la complejidad del hecho que buscan explicar.

Este tipo de género de simulaciones de datos también ha sido muy útil para explicar la evolución de la COVID-19. Además de los trackers estudiados en la sección anterior, las piezas visuales han tenido una relevancia determinante en la cobertura de la pandemia.

En marzo de 2020, a las pocas semanas de que la COVID-19 fuese declarada como pandemia mundial, *The Washington Post* publicó una simulación que demuestra cómo la distancia social puede ayudar a detener la propagación del coronavirus <sup>194</sup>.

García-Alonso indica que un buen periodismo de datos ajusta los hechos a su auténtico nivel, dando una idea real sobre su evolución <sup>195</sup>. Como ejemplo, cita un estudio sobre este diario: “Uno de los ejemplos de ese buen periodismo, que les brinda a sus lectores información relevante y de utilidad para protegerse del brote infeccioso, es el maravilloso ejercicio de periodismo de datos que hizo *The Washington Post* para ayudar a entender por qué es necesario auto aislarnos evitando así el contagio” <sup>196</sup>.

Dicha pieza, traducida a más de trece idiomas y que se convirtió en el artículo más leído del diario estadounidense en su historia, triplicando el número de visitas de la segunda historia más leída <sup>197</sup>, describe un escenario imaginario con el fin de simplificar y explicar cómo las personas contraen el coronavirus y qué pueden hacer para evitar contagiarse.

---

<sup>194</sup> Stevens, H. (14 de Marzo de 2020). *These simulations show how to flatten the coronavirus growth curve*. The Washington Post. Consultado a 25 de febrero de 2022, disponible en: <https://www.washingtonpost.com/graphics/2020/world/corona-simulator/>

<sup>195</sup> García-Alonso, P. (2020). “Claves del Periodismo Emergente en el diario The Washington Post”, pp. 85-87, con Jaspe Nieto, Javier, Capítulo en: *Periodismos Emergentes: Transformación y Revitalización del Periodismo en la Era Digital*. 2020, Universidad Complutense de Madrid, 179 págs.

<sup>196</sup> Red Ética (2020). "Coronavirus: 10 virtudes del periodismo responsable ante la pandemia", en *Fundación Gabo*, Cartagena de Indias-Colombia 17.03.2020 en <https://fundaciongabo.org/es/etica-periodistica/recursos/coronavirus-10-virtudes-del-periodismo-responsable-ante-la-pandemia>

<sup>197</sup> Kelly T. (30 de marzo de 2020). *Simulating a pandemic*. DataJournalism.com. Consultado a 26 de febrero de 2022, disponible en: <https://datajournalism.com/read/longreads/simulating-a-pandemic>

El artículo interactivo introduce al lector en un escenario imaginario, similar a la crisis de la COVID-19, pero mucho más sencillo. De esta forma, el periodista Harry Stevens, autor de la pieza, es capaz de extrapolar el concepto de distancia social y visualizar sus consecuencias en un mundo cerrado de un centenar de personas representadas por círculos de diferentes colores en función de su estado de salud.

A continuación se reproduce la descripción que se hace en la pieza de este escenario imaginario:

“Para entender por qué, podemos utilizar una simulación acerca de cómo se extendería una enfermedad ficticia entre la población. Vamos a llamar a esta enfermedad inventada ‘*simulitis*’.

Se extiende incluso más fácilmente que la COVID-19: vamos a suponer que si una persona sana entra en contacto con una persona enferma, la persona sana se contagia siempre. No hace falta mucho tiempo para que en una población de cinco personas todas se contagien de simulitis.

Como en la vida real, en algún momento estas personas se recuperan. En esta simulación, una persona ya recuperada no puede contagiar la simulitis a una persona sana, ni puede volver a enfermarse tras entrar en contacto con un enfermo.<sup>198</sup>”

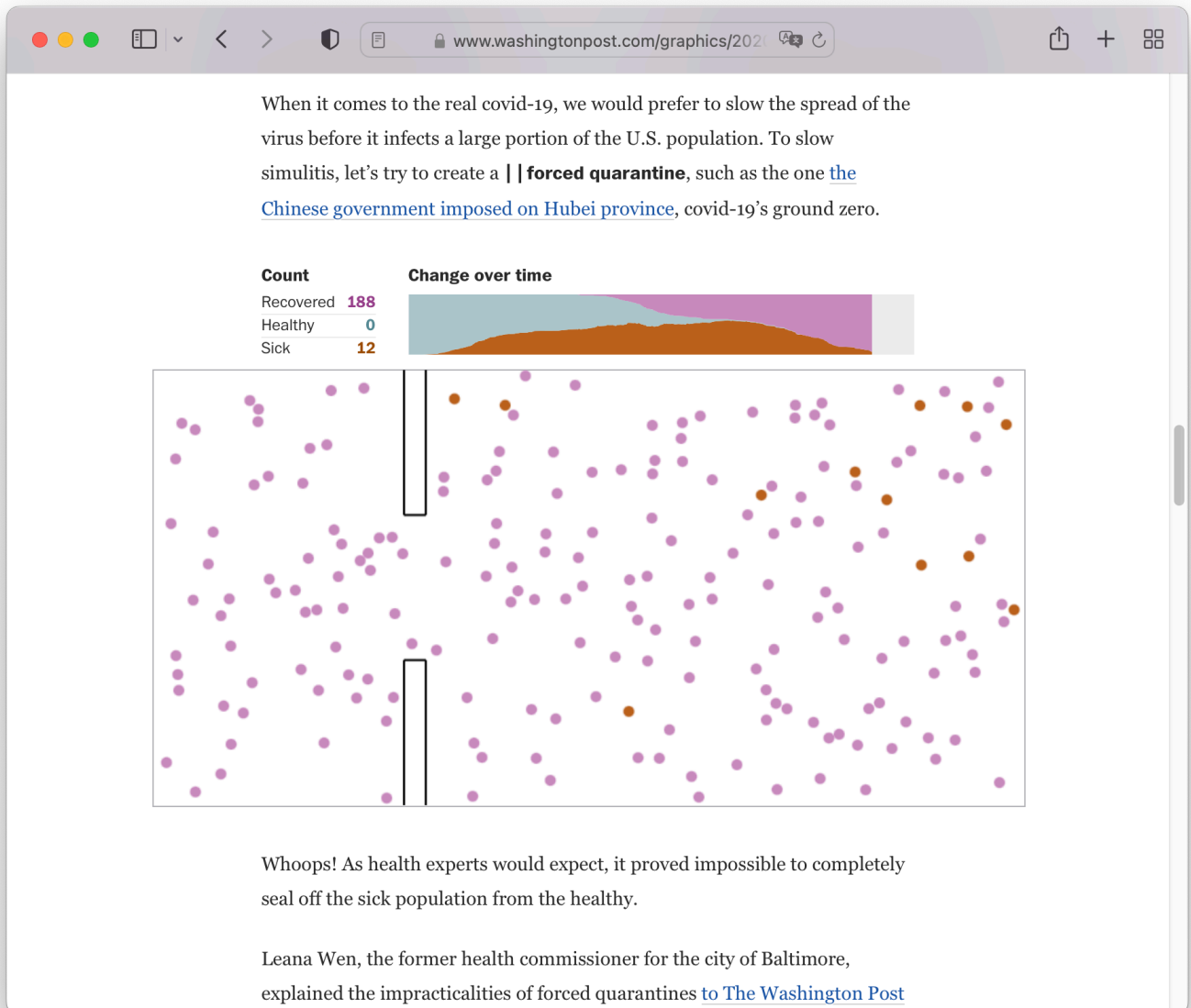
Una vez introducido el escenario, la pieza incluye una serie de simulaciones que, gracias a la automatización de procesos, se ejecutan cada vez que el lector visita la página. Estas simulaciones muestran diferentes escenarios y estados de salud de la población en función de si se ha practicado la distancia social o si se han aplicado cuarentenas.

---

<sup>198</sup> Stevens, H. (14 de Marzo de 2020). *Por qué brotes como el del coronavirus crecen exponencialmente y cómo ‘aplanar la curva*. The Washington Post. Consultado a 25 de febrero de 2022, disponible en: <https://www.washingtonpost.com/graphics/2020/world/corona-simulator-spanish/>

De esta forma, y mediante estos escenarios, la pieza presenta los efectos de la distancia social durante la crisis de la COVID-19 de forma visual y digerible para el lector.

Figura 22. Visualización de datos con la simulación de personas enfermas, sanas y recuperadas del artículo de *The Washington Post* titulado *Why outbreaks like coronavirus spread exponentially, and how to “flatten the curve”* .



Fuente: *The Washington Post*.

Aunque puede resultar imperceptible para el lector, el resultado de la simulación es diferente cada vez que la página se actualiza.

Por ende, la automatización de procesos juega un papel fundamental para poder contar este tipo de historias. La automatización está presente en el uso del lenguaje de programación (JavaScript) utilizado en la pieza, además de en las funciones y algoritmos que ejecutan las simulaciones.

Por otro lado, para explicar el fenómeno de redistricting o redistribución de distritos electoral que se da en Estados Unidos cada década en base a los cambios de población del censo, *The New York Times* creó una pieza interactiva donde el lector puede dibujar sus propios distritos <sup>199</sup>. Para ello, los periodistas del diario estadounidense plantearon un país imaginario *Hexapolis* y ofrecieron al lector el rol de legislador gracias al cual podía determinar la distribución de los distritos.

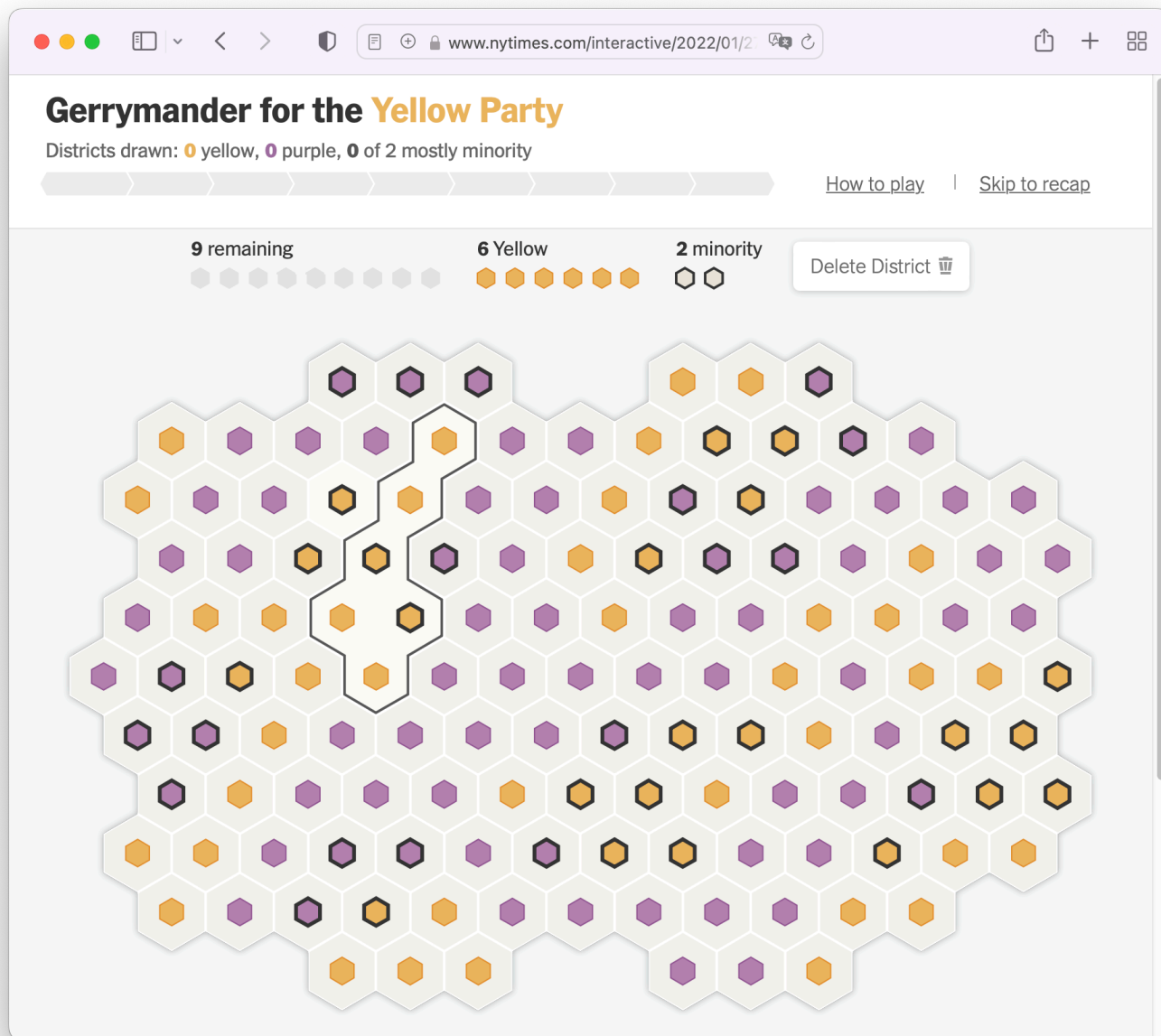
Gracias a este planteamiento, mientras interactúa con la pieza periodística, el lector puede aprender las diferentes leyes relacionadas con la redistribución de distritos electorales, las dinámicas y las estrategias que los legisladores siguen para crear y dibujar nuevos mapas y los diferentes escenarios que pueden darse en la creación de nuevos distritos.

Mediante la metáfora de *Hexapolis*, *The New York Times* es capaz de reducir un tema relativamente complejo a un juego de niños para el lector

---

<sup>199</sup> Koeze, E., Lu, D. y Smart, C. (27 de enero de 2022). *Can you gerrymander your party to power?* The New York Times. Consultado a 23 de febrero de 2022, disponible en: <https://www.nytimes.com/interactive/2022/01/27/us/politics/congressional-gerrymandering-redistricting-game-2022.html>

Figura 23. Artículo *Can you Gerrymander your party to power?* publicado por *The New York Times*.



Fuente: *The New York Times*.

## ***Newsgames* o videojuegos de noticias**

El uso de la automatización también se ve reflejado en otro género periodístico concebido como un juego para el lector con el fin de transmitir información. Los *newsgames* o juegos con vocación informativa son piezas visuales desarrolladas por los periodistas de un medio de comunicación que, mediante la aplicación de técnicas propias de los videojuegos, pretenden informar de un problema o situación al lector.

Los juegos de noticias requieren un alto grado de interacción por parte del lector y están concebidos para informar al lector a la vez que juega y avanza en las diferentes fases del juego.

Ferrari, Schweizer y Bogost ofrecen una revisión de las piezas que han surgido en la intersección entre el periodismo y los videojuegos <sup>200</sup>. Según los autores, los videojuegos simulan diferentes procesos mediante el uso de textos, imágenes, sonidos y vídeos. Los videojuegos combinan todos estos elementos para construir modelos con los que el usuario puede jugar e interactuar.

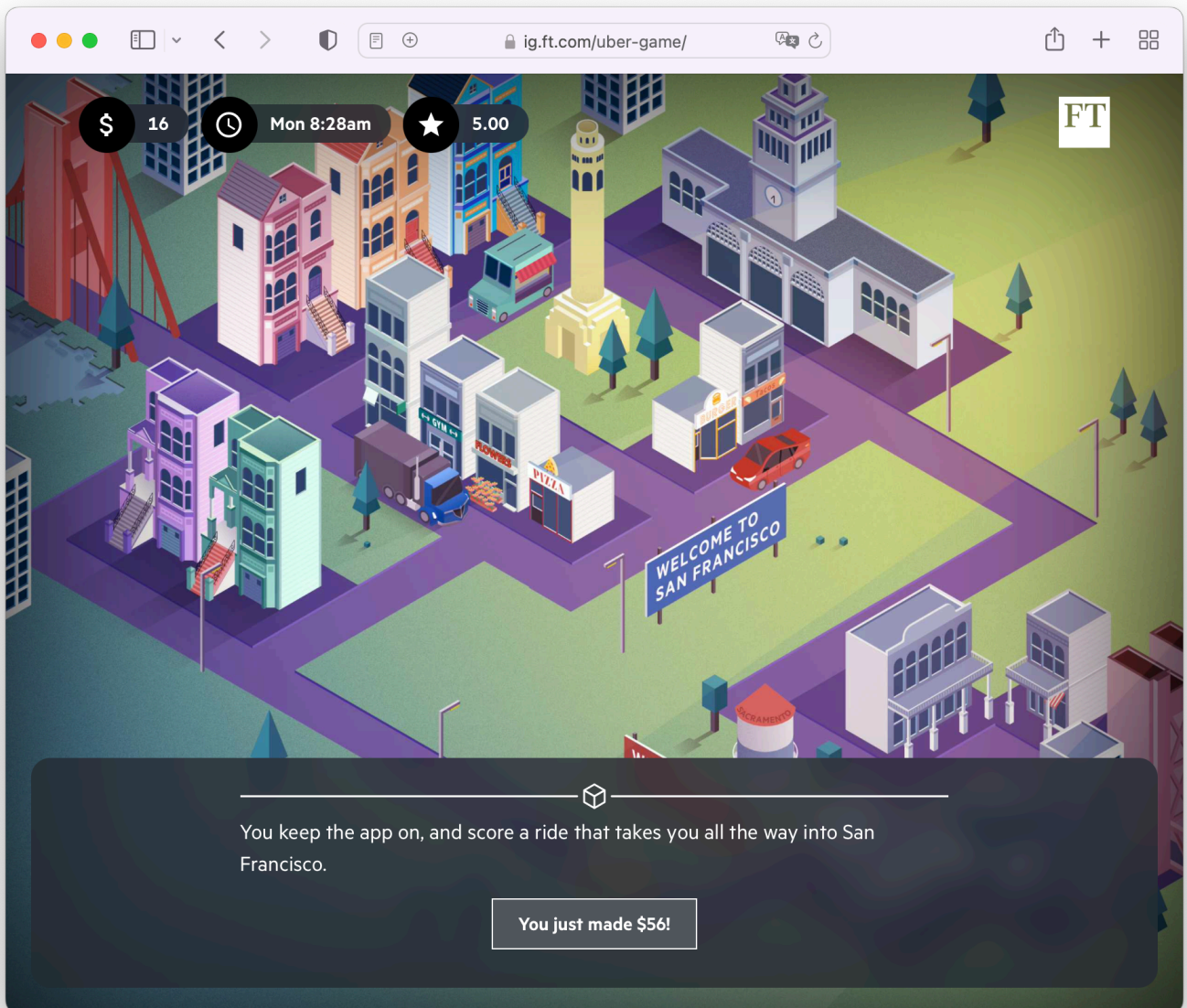
Un ejemplo de este género donde la automatización y el periodismo robot tienen un rol determinante es *The Uber Game* <sup>201</sup>, un juego publicado por *Financial Times*. El juego sitúa al lector en el rol de conductor de la compañía estadounidense de servicios de transporte, Uber. Mientras juega, el lector tendrá que tomar una serie de decisiones que repercutirán en las horas trabajadas y el dinero ganado por cada viaje. De esta forma, las decisiones que el lector tome cambiarán el resultado de la historia, acercándole a la vida diaria que lleva un trabajador de Uber.

---

<sup>200</sup> Ferrari, S., Schweizer, B., Bogost, I. (2010). *Newsgames: Journalism at Play*. United Kingdom: MIT Press, p. 6.

<sup>201</sup> Financial Times. (5 de octubre de 2017). *Could you make it as an uber driver?* Consultado a 14 de febrero de 2022, disponible en: <https://ig.ft.com/uber-game/>

Figura 24. Escena del juego *The Uber Game* desarrollado por *Financial Times*.



Fuente: Financial Times.



# **Conclusiones**



Tras examinar con detalle la evolución de la disciplina definida como periodismo robot en esta tesis y los diferentes momentos donde la automatización juega un papel fundamental en el proceso periodístico, cabe extraer una serie de puntos fundamentales a modo de conclusión.

En los capítulos de esta disertación se ha tratado de dibujar el estado actual del periodismo robot, qué representa y qué implicaciones tiene para el futuro de la disciplina periodística y de los profesionales de la información.

### **Nuevos caminos para la profesión periodística**

El periodismo robot ha abierto nuevos caminos para la profesión periodística. Gracias a la aplicación de la tecnología, los algoritmos y la automatización, disciplinas como el periodismo de investigación y el periodismo visual han crecido de manera exponencial en los últimos años.

La aplicación de estos sistemas tecnológicos, y en ocasiones inteligentes, ha permitido que los periodistas investiguen historias que hace una década no hubiesen podido realizar. De la misma forma, el periodismo visual se ha hecho un hueco en la mayoría de las redacciones de los medios de comunicación, gracias a las posibilidades que estas técnicas ofrecen al

profesional de la información para interactuar con el medio web y visualizar millones de datos. Por otro lado, la automatización de procesos ha permitido que los robots escriban noticias breves con éxito, liberando al periodista de responsabilidades menores.

Además, el periodismo robot ha facilitado la creación de nuevos géneros periodísticos como los bots, los trackers, las simulaciones o los juegos, estudiados en esta disertación. Estos son géneros con un recorrido breve en la profesión periodística pero tienen un gran potencial de evolución.

## **Desafíos**

La automatización de procesos, los algoritmos, el uso de la inteligencia artificial, en definitiva, el periodismo robot no es una técnica buena o mala. Su diseño, su concepción es apropiada o fallida. Los algoritmos, el código y los lenguajes de programación que potencian estas técnicas carecen de intenciones, son inertes. Las personas, en el caso estudiado, diseñan, desarrollan, operan y gestionan los algoritmos. Los profesionales de la información son las personas que dotan a los algoritmos de vida.

De esta forma, son los profesionales de la información quienes definen los algoritmos y les alimentan de datos para que puedan ejecutar sus funciones y ofrecer resultados. Al mismo tiempo, son las personas (aunque muchas veces no lo sepan) quienes se ven afectadas por esos algoritmos y quienes padecen las consecuencias de las decisiones positivas o negativas que toman los algoritmos <sup>202</sup>.

Diakopoulos enfatiza que, al mismo tiempo que las capacidades de los algoritmos mejoran los resultados de disciplinas como la minería de datos y la curación y creación de contenidos, los periodistas y la sociedad no deben olvidar el rol que las personas juegan en los usos y aplicaciones futuras de los algoritmos <sup>203</sup>.

---

<sup>202</sup> Diakopoulos, N. (2019). *Automating the News: How Algorithms Are Rewriting the Media*. United Kingdom: Harvard University Press, p. 240.

<sup>203</sup> Ibid, p. 241.

En ocasiones, el papel y la responsabilidad del periodista en la creación de un algoritmo será directo, si escribe el código que potencia un bot de noticias. Sin embargo, no siempre la implicación de las personas es tan evidente. Una persona puede influir en el resultado de un algoritmo de forma indirecta mediante la selección y preparación de los datos para entrenar a un algoritmo.

En definitiva, las personas, en este caso, los periodistas, tienen la última palabra y la responsabilidad sobre cómo estos sistemas y algoritmos operan en el entorno y qué influencia tienen tanto la sociedad como en los medios de comunicación. Por ello, Diakopoulos propone que el futuro de los algoritmos tiene que pasar y centrarse en las personas y en el impacto que tienen sobre ellas.

De acuerdo con este autor, el objetivo principal es que el periodista pueda controlar el poder de los algoritmos a la hora de personalizar contenidos y aplicarlos a nuevas tareas periodísticas, como el análisis o la visualización de datos <sup>204</sup>. Bajo el punto de vista del autor, el objetivo principal es diseñar estos algoritmos, para que no sólo sean beneficiosos desde el punto de vista puramente económico, sino también para las personas y la sociedad. Los periodistas y otros diseñadores de algoritmos deben determinar los valores de estos algoritmos de forma deliberada y abierta, con el fin de que se puedan auditar si es necesario.

Además, la automatización también genera dudas sobre qué implica codificar el juicio periodístico y editorial en un algoritmo <sup>205</sup>. Se trata de cuestiones que deben ser revisadas con el tiempo y la ejecución de estas técnicas por parte de los medios de comunicación.

---

<sup>205</sup> Lecompte, C. (19 de octubre de 2020). *Automation in the newsroom*. Nieman Reports. Consultado a 4 de marzo de 2022, disponible en: <https://niemanreports.org/articles/automation-in-the-newsroom/>

## **Mejora de las condiciones laborales**

La introducción de la automatización y el uso de algoritmos no va en detrimento de los profesionales de la información. El objetivo final es ayudar al periodista o editor a completar y complementar su trabajo.

Para la mayoría de las organizaciones mediáticas, la automatización de parte del proceso periodístico y de publicación de contenidos es una forma de reducir la carga de trabajo de sus periodistas, además de sacar el máximo partido a los nuevos recursos tecnológicos y los datos disponibles.

Según los ejemplos descritos a lo largo de esta tesis, la automatización de ciertas tareas del proceso periodístico y de publicación de historias, reduce la carga de trabajo de los periodistas. De esta forma, el periodista puede centrarse en otros aspectos más relevantes de su trabajo, como invertir más tiempo y esfuerzo en investigaciones de largo recorrido o en explorar otro tipo de datos, ideas o recursos durante su jornada laboral <sup>206</sup>.

El periodismo robot reemplaza los procesos repetitivos que el periodista tiene que realizar en su día a día. De esta forma, el profesional de la información puede invertir su tiempo en tareas más productivas. Por último, la automatización de tareas repetitivas permite ejecutar acciones o recopilar datos, que de otra forma no podrían realizarse, debido a la capacidad de concentración que requieren y a la cantidad de tiempo que implican.

## **El periodismo robot no reemplazará a los periodistas**

Las recientes pérdidas de puestos laborales en el sector del periodismo y la lenta desaparición de cientos de proyectos de periodismo local por falta de recursos no tienen una relación directa con la incorporación de la tecnología en el proceso periodístico. El desarrollo

---

<sup>206</sup> Ibid.

de la automatización de forma controlada en las labores periodísticas no implica la pérdida o la disminución de trabajos.

Este matrimonio entre tecnología y periodismo sí implica la aparición de un nuevo perfil profesional: un periodista especializado en el uso de datos y los lenguajes de programación para producir su trabajo. Estos conocimientos implican una barrera de entrada al rol, a la vez que ofrecen diferentes oportunidades tanto a estudiantes de la profesión periodística como a profesionales.

### **En continuo desarrollo y crecimiento**

La intersección entre la tecnología y el periodismo está en auge. Este crecimiento puede observarse en el número de piezas periodísticas publicadas, donde la automatización tiene un papel fundamental. Este auge del periodismo robot también puede verse en el crecimiento de los medios de comunicación con más recursos y en las áreas donde están enfocando sus esfuerzos. Éstas son la tecnología, el análisis de datos y la inteligencia artificial.

Los ingenieros son el segundo grupo más grande de trabajadores en *The New York Times*, después de los periodistas. Éste es un hecho digno de destacar, ya que, si se elimina a los periodistas de la ecuación, base de cualquier proyecto de la información, los ingenieros y los perfiles tecnológicos son el grupo más grande dentro de la masa laboral del diario <sup>207</sup>.

Por su parte, la apuesta por la tecnología de *The Washington Post* muestra también muestra hacia dónde se dirige el sector. Desde que el fundador de Amazon, Jeff Bezos, adquiriese la publicación en 2013, el diario norteamericano inició una estrategia de

---

<sup>207</sup> Doctor, K. (30 de Julio de 2020). *The New York Times' new CEO, Meredith Levien, on building a world-class digital media business - and a Tech Company*. Nieman Lab. Consultado a 2 de agosto de 2020, disponible en: <https://www.niemanlab.org/2020/07/newsonomics-the-new-york-times-new-ceo-meredith-levien-on-building-a-world-class-digital-media-business-and-a-tech-company/>

digitalización y modernización tecnológica que ha sido dotada de más fondos y relevancia en los últimos años.

Uno de los pilares clave de esta estrategia de *The Washington Post* es *Arc XP* <sup>208</sup>: una división del diario que se ha establecido como uno de los líderes de referencia en el mercado de las plataformas de publicación digitales o Content Management System (CMS).

Según datos del propio diario, la plataforma creció a pasos agigantados en 2021 y ya es la plataforma de publicación de 1.900 páginas web por todo el mundo, llegando a más de 1.500 millones de usuarios únicos al mes. Tras siete años en el mercado, Arc XP ofrece sus servicios a numerosas marcas y empresas de diferentes sectores e industrias, además de editoras y emisoras <sup>209</sup>.

De acuerdo con el propio diario, este crecimiento es el resultado de “una evolución estratégica y de innovaciones continuas” <sup>210</sup>. Se trata de una estrategia que muestra la apuesta del sector por la tecnología y la automatización de procesos, no sólo para mejorar el trabajo de los redactores sino también para mejorar económicamente.

### **La pandemia ha acelerado el uso de la automatización**

Uno de los factores que ha contribuido al crecimiento acelerado del periodismo robot ha sido la pandemia mundial por la COVID-19. La crisis sanitaria ha acelerado el uso de la tecnología y de la automatización en el apartado de contenidos, con la meta de ofrecer información actualizada al minuto y precisa al lector. Numerosos ejemplos muestran cómo las piezas visuales publicadas por *The New York Times*, *The Washington Post* o *South China*

---

<sup>208</sup> WashPostPR. (16 de diciembre de 2021). *Arc XP in 2021*. The Washington Post. Consultado a 12 de enero de 2022, disponible en: <https://www.washingtonpost.com/pr/2021/12/16/arc-xp-2021/>

<sup>209</sup> Ibid.

<sup>210</sup> Ibid.

*Morning Post*, por recordar algunos ejemplos, se han beneficiado de la automatización de procesos para reaccionar más rápido a la actualidad.

Los resultados de la encuesta sobre el estado del periodismo de datos en 2021, realizada por la organización European Journalism Centre (EJC), también incide en el papel que la pandemia ha tenido en el periodismo de datos y, por ende, en la adopción de diversos procesos tecnológicos en la labor periodística como la automatización. De acuerdo con las conclusiones del estudio, “el impacto de la pandemia en el campo del periodismo de datos es claro”<sup>211</sup>.

En el análisis de las más de 1.200 respuestas recibidas por periodistas de diferentes partes del mundo, European Journalism Centre concluye cómo la crisis de salud global ha cambiado el curso del periodismo de datos en las redacciones.

Un 25% de los participantes afirma que se han introducido en el periodismo de datos debido a la pandemia. Además, un 46% de los entrevistados concluye que la crisis sanitaria ha reforzado la labor y la misión del periodismo de datos, además de haber incrementado la alfabetización y conocimientos de las audiencias alrededor del mundo de los datos.

## **Nuevos retos para el periodista y los medios de comunicación**

La tecnología avanza más rápido que el periodismo. Los rápidos avances a nivel tecnológico presentan una serie de desafíos para el periodista. El primero de los desafíos es una cuestión económica. Los medios de comunicación con mayores recursos han sido capaces de incorporar profesionales con los conocimientos necesarios para desarrollar análisis complejos de enormes bases de datos, automatizar procesos repetitivos o investigar algunos de los algoritmos con mayor influencia en la sociedad.

---

<sup>211</sup> Abellán, A. (16 de febrero de 2022). *State of Data Journalism Survey 2021: 11 surprising findings*. DataJournalism.com. Consultado a 5 de marzo de 2022, disponible en: <https://datajournalism.com/read/blog/data-journalism-survey-2021#pandemic>

Por el contrario, los medios más pequeños han sido los que más han sufrido a la hora de dar este salto. La falta de recursos, incluso para cubrir las bases del periodismo local, provoca la falta de profundidad en el tipo de historias que pueden ofrecer o en los productos informativos que pueden poner a disposición de sus lectores.

Las diferentes velocidades a las que se mueve el periodismo y la tecnología, supone además un reto para los profesionales de la información. Para poder trabajar y manipular una base de datos, que es el elemento más básico del periodismo robot, se requiere una serie de conocimientos técnicos para los que un buen número de profesionales de la información experimentados no están preparados.

Por el contrario, en los últimos años ha surgido un perfil profesional multidisciplinar. Esta nueva figura corresponde al periodista con conocimientos computacionales, estadísticos y de ciencia de datos. Pese a la amplitud de esos conocimientos que incluso tocan otros campos de conocimiento alejados del periodismo, el objetivo del profesional de la información es el mismo: elevar el producto periodístico final, contribuyendo a descubrir y contar nuevas historias.

Es complicado atraer a otros profesionales externos, como pueden ser ingenieros o estadísticos, al sector del periodismo. Esto se debe, fundamentalmente, a dos razones principales. En primer lugar, debido a la mayor competencia que existe a nivel de salarios en otros sectores. Éste es un hecho que influye negativamente en la capacidad de captación de este tipo de profesionales en el sector del periodismo.

En segundo lugar, el mundo del periodismo requiere un compromiso y un interés especial, debido al tipo de trabajo que supone. Este compromiso hace que la atracción y retención de otros perfiles profesionales sea todavía menor en comparación con otros sectores.

De esta forma, los presupuestos limitados de los medios y la escasez de profesionales y recursos en el sector contribuyen a la diferencia de velocidades en el progreso del

periodismo con respecto a la tecnología. Aunque los medios de comunicación más relevantes a nivel mundial como *The New York Times* o *The Washington Post* han sido capaces de reducir esa distancia entre periodismo y tecnología, los medios de tamaño medio o pequeño y con presupuestos apretados son los que más sufren estados de velocidades por la cantidad de historias que no pueden cubrir, debido a la falta de recursos, tanto profesionales como económicos.

## **Investigaciones futuras**

Este rápido crecimiento y adopción de la automatización en el sector del periodismo también implica numerosos desafíos, además de nuevos campos a explorar e investigar.

El uso de la automatización en periodismo es un campo relativamente nuevo, pero muy amplio por todas las ramificaciones y posibilidades que presenta, como se ha expuesto a lo largo de esta tesis. Debido a este hecho, se han fundado múltiples disciplinas y conceptualizaciones para describir los procesos tecnológicos de los que un periodista puede servirse para realizar un mejor trabajo o más eficiente.

Esta tesis presenta el término ‘periodismo robot’ como paraguas conceptual para estudiar todos los términos y disciplinas que los académicos en el campo del periodismo han acuñado en los últimos años. Pese a ello, no existe consenso a la hora de referirse a la especialización o técnicas que involucran el uso de datos, automatización, algoritmos y lenguajes de programación.

Por ello, esta investigación puede extenderse con el fin de presentar un marco más extenso para el estudio de la automatización y la tecnología aplicada al periodismo, definiendo de forma más específica y más amplia roles, técnicas y características de cada una de las tareas habituales que un periodista realiza en su día a día.

Esta tesis se centra primordialmente en el uso y aplicación de la automatización, los robots y la tecnología en las técnicas de investigación y el lenguaje visual. Aunque en los

diferentes capítulos de esta tesis se hace referencia a otros formatos como el vídeo o el audio, hay espacio para profundizar y estudiar la relación de estos formatos con la tecnología.

En definitiva, la última y más relevante conclusión que puede extraerse de esta disertación es que, pese a que la tecnología refuerza y amplifica las habilidades del periodista, el factor humano, su juicio y visión editorial, son claves a la hora de ejecutar los procesos tecnológicos analizados en esta tesis.

Los robots, la tecnología y la inteligencia artificial mejoran las capacidades del profesional de la información para identificar contenidos relevantes, crear informaciones o analizar datos, entre otros, pero en ningún caso pueden sustituir a la figura del periodista. Para el correcto funcionamiento y evolución de la disciplina presentada en esta tesis, la supervisión y toma de decisiones del periodista es clave en el presente y el futuro del periodismo robot.



**REFERENCIAS  
BIBLIOGRÁFICAS**



Abellán, A. (16 de febrero de 2022). State of Data Journalism Survey 2021: 11 surprising findings. DataJournalism.com. Consultado a 5 de marzo de 2022, disponible en: <https://datajournalism.com/read/blog/data-journalism-survey-2021#pandemic>

Alfred, R. (4 de noviembre de 2010). Nov. 4, 1952: Univac gets election right, but CBS balks. Wired. Consultado a 5 de febrero de 2022, disponible en: <https://www.wired.com/2010/11/1104cbs-tv-univac-election/>

Algorithm Watch. (2020). Automating Society Report 2020, Algorithm Watch. Disponible en: <https://automatingsociety.algorithmwatch.org/wp-content/uploads/2020/12/Automating-Society-Report-2020.pdf>

Allsop, J. (17 de marzo de 2021). The COVID Tracking Project is (nearly) gone. Can we see clearly now, Columbia Journalism Review (CJR). Consultado a 21 de enero de 2022, disponible en: [https://www.cjr.org/the\\_media\\_today/the-covid-tracking-project-is-nearly-gone-can-we-see-clearly-now.php](https://www.cjr.org/the_media_today/the-covid-tracking-project-is-nearly-gone-can-we-see-clearly-now.php)

Angwin, J. (25 de febrero de 2020). A letter from the editor – The Markup. The Markup. Consultado a 20 de marzo de 2021, disponible en: <https://themarkup.org/2020/02/25/editor-letter-julia-angwin>

Baruch, J., Aubert, R., Ferrer, M., Sanchez, L., Dagorn, G. y Breteau, P. (s.f.). Covid-19 : Le Tableau De Bord de l'épidémie. Le Monde. Consultado a 8 de marzo de 2022, disponible en: [https://www.lemonde.fr/les-decodeurs/article/2022/06/09/covid-19-le-tableau-de-bord-de-l-epidemie\\_6038751\\_4355773.html](https://www.lemonde.fr/les-decodeurs/article/2022/06/09/covid-19-le-tableau-de-bord-de-l-epidemie_6038751_4355773.html)

Bhatia, G., Dutta, P. K. y McClure, J. (s.f.). Covid-19: The Latest Global Statistics, charts and maps. Reuters. Consultado a 8 de marzo de 2022, disponible en: <https://graphics.reuters.com/world-coronavirus-tracker-and-maps/>

Benton, J. (4 de agosto de 2016). Sensor journalism may have lost some of its buzz, but it's also gotten cheaper and easier to pull off. Nieman Lab. Consultado a 3 de junio de 2019, disponible en: <https://www.niemanlab.org/2021/08/sensor-journalism-may-have-lost-some-of-its-buzz-but-its-also-gotten-cheaper-and-easier-to-pull-off/>

Berkowitz, B. D. (2018). *Playfair: The True Story of the British Secret Agent who Changed how We See the World*. United States: George Mason University Press.

Blanco, A. y Galocha, A. (24 de mayo de 2022). How election modeling can help us understand who might win. The Washington Post. Consultado a 24 de mayo de 2022, disponible en: <https://www.washingtonpost.com/elections/interactive/2022/how-election-models-work/>

Blanco, A. y Sauras, J. (1 de junio de 2021). Drawing the curve: Data visualization and COVID-19 A pandemic year in graphics. Tow Center for Digital Journalism at Columbia's Graduate School of Journalism. Consultado a 2 de junio de 2021, disponible en: <https://mailchi.mp/95bdcee6e0d7/weekly-updates-on-covid-19s-impact-on-journalism-2083305?e=e8e508e35f>

Blankespoor, E., Dehaan, E. y Zhu, C. (10 de mayo de 2017). Robo-journalism and capital markets. Stanford Graduate School of Business. Consultado a 5 de abril de 2021, disponible en: <https://www.gsb.stanford.edu/faculty-research/working-papers/robo-journalism-capital-markets>

Bounegru, L. y Grey, J. (Eds.). (2012). *The Data Journalism Handbook: Towards A Critical Data Practice*. Amsterdam University Press, p. 237. Disponible en: 10.1017/9789048542079

Bradshaw, P. (2013). *Scraping for Journalists: How To Grab Data From Hundreds of Sources, Put It In a Form You Can Interrogate and Still Hit Deadlines*. Canada: Leanpub.

Bronner, L. (2 de noviembre de 2021). How The Washington Post will model possible outcomes in the Virginia governor's race. The Washington Post. Consultado a 11 de enero de 2022, disponible en: <https://www.washingtonpost.com/elections/2021/11/02/election-model-explained/>

Cairo, Alberto (2016). *The Truthful Art: Data, Charts, and Maps for Communication*. Berkeley, CA: Peachpit Press.

Cappiello, D. (29 Julio de 2011). Chronicle cross-county study reveals risky load of 'air toxics'. Chron. Consultado a 6 de junio de 2020, disponible en: <https://www.chron.com/news/article/Chronicle-cross-county-study-reveals-risky-load-1643020.php>

Carrasco Polaino, R. y Viviar Flores, J. (2020). Periodismo de datos y visualización en Periodismo y nuevos medios: Perspectivas y retos, pp. 249–273. Editorial GEDISA.

CDC. (11 de mayo de 2022). Vital statistics rapid release - Provisional drug overdose data. Centers for Disease Control and Prevention. Consultado a 2 de enero de 2022, disponible en: <https://www.cdc.gov/nchs/nvss/vsrr/drug-overdose-data.htm>

CDC. (17 de noviembre de 2021). Drug overdose deaths in the U.S. top 100,000 annually. Centers for Disease Control and Prevention. Consultado a 2 de enero de 2022, disponible en: [https://www.cdc.gov/nchs/pressroom/nchs\\_press\\_releases/2021/20211117.htm](https://www.cdc.gov/nchs/pressroom/nchs_press_releases/2021/20211117.htm)

Chaparro Domínguez, M.A. (2013). La evolución del periodismo de precisión: el blog de The Guardian sobre periodismo de datos, en Actas In Congreso Internacional de la Comunicación y Sociedad Digital, Universidad Internacional de La Rioja, Logroño, 18 y 19 de abril de 2013. Disponible en: <https://dialnet.unirioja.es/servlet/articulo?codigo=4247831>

Chellel, K., Mokwa, W., Lin, J. C. F. y Randolph, S. (29 de marzo de 2022). A Plastic Bag's 2,000-Mile Journey Shows the Messy Truth About Recycling. Bloomberg. Consultado a 12 de abril de 2022, disponible en: <https://www.bloomberg.com/graphics/2022-tesco-recycle-plastic-waste-pledge-falls-short/>

Cherian, J. y Bronner, L. (2020). How The Washington Post Estimates Outstanding Votes for the 2020 Presidential Election. Washington Post Engineering. Disponible en: [https://s3.us-east-1.amazonaws.com/elex-models-prod/2020-general/write-up/election\\_model\\_writeup.pdf](https://s3.us-east-1.amazonaws.com/elex-models-prod/2020-general/write-up/election_model_writeup.pdf)

CNN (16 de marzo de 2021). America wasn't ready for Covid-19. These newsrooms helped guide the way, CNN. Consultado a 21 de enero de 2022, disponible en: <https://www.cnn.com/2021/03/16/media/the-atlantic-stat-pandemic-journalism/index.html>

Coddington, Mark (2015) Clarifying Journalism's Quantitative Turn, *Digital Journalism*, 3:3, 331-348.

Cohen, S. y Bruzese, L. (2001). Numbers in the Newsroom: Using Math and Statistics in News. United States: Investigative Reporters and Editors, Incorporated.

Confessore, N., Dance, G., Harris, R. y Hansen, M. (27 de enero de 2018). The follower factory. The New York Times. Consultado a 1 de junio de 2019, disponible en: <https://www.nytimes.com/interactive/2018/01/27/technology/social-media-bots.html>

Cox, M. (2000). The Development of Computer-Assisted Reporting. Estudio presentado en el coloquio Newspaper Division, Association for Education in Journalism and Mass Communication, Southeast Colloquium el 17 de Marzo de 2000 en la University de Carolina del Norte, Chapel Hill.

CSSEGISandData. (s.f.). CSSEGISANDDATA/covid-19: Novel coronavirus (COVID-19) cases, provided by JHU CSSE. GitHub. Consultado a 21 de enero de 2022, disponible en: <https://github.com/CSSEGISandData/COVID-19>

Dader, J.L. y Gómez Fernández, P. (1993). Periodismo de precisión: Una nueva metodología para transformar el periodismo, en *Articles, análisis* 15, pp. 99-116.

Danzon-Chambaud, S. (17 de septiembre de 2021). How news organizations used Automated News to cover COVID-19. Poynter. Consultado a 13 de enero de 2022, disponible en: <https://www.poynter.org/tech-tools/2021/how-news-organizations-used-automated-news-to-cover-covid-19/>

Danzon-Chambaud, S. (6 de agosto de 2021). Covering covid-19 with Automated News. *Columbia Journalism Review*. Consultado a 14 de enero de 2022, disponible en: [https://www.cjr.org/tow\\_center\\_reports/covering-covid-automated-news.php](https://www.cjr.org/tow_center_reports/covering-covid-automated-news.php)

Data Visualization Society. (2021). State of the industry survey — 2021 report. Data Visualization Society. Disponible en el portal: <https://www.datavisualizationsociety.org/report-2021> y en un formato .pdf en: [https://static1.squarespace.com/static/5c6055b5fb18206d45d6b27e/t/61f0573657c6471e4d53dc58/1643140919414/DVS\\_State+of+the+Industry+2021+Report.pdf](https://static1.squarespace.com/static/5c6055b5fb18206d45d6b27e/t/61f0573657c6471e4d53dc58/1643140919414/DVS_State+of+the+Industry+2021+Report.pdf)

Datajournalism.com. (s.f.). Data scraping for stories, *Datajournalism.com*. Consultado a 12 de septiembre de 2021, disponible en: <https://datajournalism.com/read/newsletters/data-scraping-for-stories>

Department of Computer Science at the University of Helsinki. (4 de Julio de 2017). Valtteri the Election Bot generates news in three languages. *University of Helsinki*. Consultado a 15 de septiembre de 2021, disponible en: <https://www.cs.helsinki.fi/en/news/86516>

Diakopoulos, N. (15 de abril de 2020). The ethics of Predictive Journalism. *Columbia Journalism Review*. Consultado a 5 de abril de 2021, disponible en: [https://www.cjr.org/tow\\_center/predictive-journalism-artificial-intelligence-ethics.php](https://www.cjr.org/tow_center/predictive-journalism-artificial-intelligence-ethics.php)

Diakopoulos, N. (2010). A Functional Roadmap for Innovation in Computational Journalism. School of Communication and Information, Rutgers University.

Diakopoulos, N. (2019). Automating the News: How Algorithms Are Rewriting the Media. United Kingdom: Harvard University Press.

Doctor, K. (30 de Julio de 2020). The New York Times' new CEO, Meredith Levien, on building a world-class digital media business - and a Tech Company. Nieman Lab. Consultado a 2 de agosto de 2020, disponible en: <https://www.niemanlab.org/2020/07/newsonomics-the-new-york-times-new-ceo-meredith-levien-on-building-a-world-class-digital-media-business-and-a-tech-company/>

El País. (s.f.). El Mapa del coronavirus en el mundo: Así avanzan Los Contagios y las muertes día a día. EL PAÍS. Consultado a 8 de marzo de 2022, disponible en: <https://elpais.com/especiales/coronavirus-covid-19/el-mapa-del-coronavirus-en-el-mundo/>

Fahn, J. (s.f.). Feel me?, Columbia Journalism Review. Consultado a 5 de abril de 2022, disponible en: [https://archives.cjr.org/feature/feel\\_me.php](https://archives.cjr.org/feature/feel_me.php)

Fehr, T. y Williams, J. (24 de junio de 2021). 10 million data requests: How our Covid team tracked the pandemic. The New York Times. Consultado a 13 de febrero de 2022, disponible en: <https://www.nytimes.com/2021/06/24/insider/covid-tracking-data.html>

Ferrari, S., Schweizer, B., Bogost, I. (2010). Newsgames: Journalism at Play. United Kingdom: MIT Press

Financial Times. (5 de octubre de 2017). Could you make it as an uber driver? Consultado a 14 de febrero de 2022, disponible en: <https://ig.ft.com/uber-game/>

Fink, K. y Anderson, C.W. (2015). Data Journalism in the United States. Journalism Studies, 16 (4). pp. 467-481.

Fischer, S. (18 de enero de 2022). The markup doubles down on data tools as revenue grows. Axios. Consultado a 12 de enero de 2022, disponible en: <https://www.axios.com/the-markup-nonprofit-tech-news-2021-revenue-9bb7a79e-afca-4b30-a791-1deed46b1eb4.html>

Flores Vivar, J. y Salinas Aguilar, C. (2018). Algorithms, applications and Big Data, new paradigms in the process of communication and teaching-learning of data journalism. En *Revista de Comunicación*, Volumen 17, número 2.

Forrest, J. y Wahid, R. (7 de noviembre de 2021). Interview with the South China Morning Post Infographics team, Nightingale. *Nightingale Journal*. Consultado a 15 de diciembre de 2021, disponible en: <https://nightingaledvs.com/on-the-success-of-the-south-china-morning-post-infographics-team/>

French, A. y Camberg, N. (22 de abril de 2021). Inside the Covid Tracking Project's volunteer organization. *The COVID Tracking Project*. Consultado a 23 de febrero de 2022, disponible en: <https://covidtracking.com/analysis-updates/inside-the-covid-tracking-projects-volunteer-organization>

García-Alonso, P. (2012). “Empresa Informativa en 2012. Nuevo modelo de negocio”, pp. 1-9, Capítulo en *Actas IV Congreso Internacional de Cyberperiodismo y Web 2.0.pdf* (8.086Mb) 2012, Ed. Servicio editorial de Universidad del País Vasco (UPV/EHU): <https://dialnet.unirioja.es/servlet/articulo?codigo=7739731>

García-Alonso, P. (2018). “Nuevo modelo de negocio para la empresa informativa”, pp. 81-94, Capítulo en: “Perfiles actuales en la información y en los informadores”, Ed. Tecnos-Grupo Anaya, Madrid 2018, 1.ª edición, 2017, 497 págs. [https://repository.udem.edu.co/bitstream/handle/11407/5261/Perfiles\\_actuales\\_informacion\\_informadores.pdf?sequence=1&isAllowed=y](https://repository.udem.edu.co/bitstream/handle/11407/5261/Perfiles_actuales_informacion_informadores.pdf?sequence=1&isAllowed=y) Edición digital: [https://www.researchgate.net/profile/Joaquin\\_Marques2/publication/324596712\\_LOS\\_DIARIOS\\_ECONOMICOS\\_ESPAÑOLES\\_entre\\_dos\\_siglos\\_1978-2016/](https://www.researchgate.net/profile/Joaquin_Marques2/publication/324596712_LOS_DIARIOS_ECONOMICOS_ESPAÑOLES_entre_dos_siglos_1978-2016/)

links/5ad781320f7e9b28593ac24b/LOS-DIARIOS-ECONOMICOS-ESPANOLES-entre-dos-siglos-1978-2016.pdf?origin=publication\_detail

García-Alonso, P. (2020). “Claves del Periodismo Emergente en el diario The Washington Post”, pp. 85-87, con Jaspe Nieto, Javier, Capítulo en: Periodismos Emergentes: Transformación y Revitalización del Periodismo en la Era Digital. 2020, Universidad Complutense de Madrid, 179 págs.

Galindo Arranz, F. (2004). Propuesta de periodización histórica y evolución conceptual del Periodismo de Precisión, Universidad de Santiago de Compostela. Disponible en [http://webs.ucm.es/info/emp/Numer\\_10/Sum/3-04.pdf](http://webs.ucm.es/info/emp/Numer_10/Sum/3-04.pdf)

Gilmour, J. (28 de mayo de 2021). Analysis & Updates: 20,000 hours of data entry: Why we didn't Automate Our Data Collection. The COVID Tracking Project. Consultado a 16 de enero de 2022, disponible en: <https://covidtracking.com/analysis-updates/why-we-didnt-automate-our-data-collection>

Goff L. (16 de febrero de 1999). Univac predicts winner of 1952 election. CNN. Consultado a 6 de febrero de 2022, disponible en: <http://www.cnn.com/TECH/computing/9904/30/1952.idg/>

Graefe, A. (2016,). Guide to automated journalism. Columbia Journalism Review. Consultado a 4 de septiembre de 2021, disponible en: [https://www.cjr.org/tow\\_center\\_reports/guide\\_to\\_automated\\_journalism.php](https://www.cjr.org/tow_center_reports/guide_to_automated_journalism.php)

Greenway, R. (20 de abril de 2020). Tools for Public Service journalism: Q&A with Casey Miller. RJI. Consultado a 6 de enero de 2022, disponible en: <https://rjionline.org/news/tools-for-public-service-journalism-qa-with-casey-miller/>

Guardian News and Media. (10 de marzo de 2009). The Guardian Datablog. The Guardian. Consultado a 1 de mayo de 2021, disponible en: <https://www.theguardian.com/news/datablog/2009/mar/10/blogpost1>

Hamilton, J. T. y Turner, F. (2009). Accountability through Algorithm: Developing the Field of Computational Journalism. Center for Advanced Study in the Behavioral Sciences Summer Workshop. Disponible en: <http://web.stanford.edu/~fturner/Hamilton%20Turner%20Acc%20by%20Alg%20Final.pdf>

Hansen, M., Roca-Sales, M., Keegan, J.M. y King, G. (2017). Artificial Intelligence: Practice and Implications for Journalism. Tow Center for Digital Journalism y Brown Institute for Media Innovation. Disponible en: <https://academiccommons.columbia.edu/doi/10.7916/D8X92PRD>

Harvard T.H. Chan School Of Public Health. (14 de julio de 2020). Data animation shows time lag between COVID-19 cases and deaths. Harvard T.H. Chan School of Public Health. Consultado a 10 de febrero de 2022, disponible en: <https://www.hsph.harvard.edu/news/hsph-in-the-news/data-animation-shows-time-lag-between-covid-19-cases-and-deaths/>

Herzog, D. L. (2015). Data Literacy: A User's Guide. United States: SAGE Publications.

Hildén, J., Koponen, J. (2019). Data Visualization Handbook. Finland: Aalto University.

Houston, B. (13 de diciembre de 2021). The history of Data Journalism. DataJournalism.com. Consultado a 1 de mayo de 2022, disponible en: <https://datajournalism.com/read/longreads/the-history-of-data-journalism>

Houston, B. (2018). Data for Journalists: A Practical Guide for Computer-Assisted Reporting. United Kingdom: Taylor y Francis.

Howard, A.B. (2014). The Art and Science of Data-Driven Journalism. Tow Center for Digital Journalism. Consultado a 16 de Enero de 2022, disponible en: <https://academiccommons.columbia.edu/doi/10.7916/D8Q531V1>

Hullman, J. (2020). Why Authors Don't Visualize Uncertainty, IEEE Transactions on Visualization and Computer Graphics, 26, 130-139, disponible en: <https://ieeexplore.ieee.org/document/8805422>

Immersive Automation. (2018). Immersive Automation: summary of project outcomes. Immersive automation. Consultado a 15 de septiembre de 2021, disponible en: <http://immersiveautomation.com/wp-content/uploads/2018/10/Immersive-Automation-for-web-page.pdf>

IRE. (s.f.). About ire. Investigative Reporters and Editors. Consultado a 17 de enero de 2022, disponible en: <http://www.ire.org/about-ire/>

Jaspin, E. G. y Johnson, M.M. (1 de Marzo de 1987). R.I., system fails to fully check driving records of bus applicants, Providence Sunday Journal.

Keefe, J. (2016). Family Projects for Smart Objects: Tabletop Projects That Respond to Your World. United States: Make Community, LLC.

Keegan, J. y Lecher, C. (24 de marzo de 2021). We ran tests on every state's COVID-19 vaccine website – the markup. We Ran Tests on Every State's COVID-19 Vaccine Website – The Markup. Consultado a 12 de enero de 2022, disponible en: <https://themarkup.org/blacklight/2021/03/24/we-ran-tests-on-every-states-covid-19-vaccine-website>

Kelly T. (30 de marzo de 2020). Simulating a pandemic. DataJournalism.com. Consultado a 26 de febrero de 2022, disponible en: <https://datajournalism.com/read/longreads/simulating-a-pandemic>

Kelly, T., Houston, B. y Doig, S. (13 de diciembre de 2021.). Episode 42: Conversation with Brant Houston and Stephen Doig. SoundCloud. Consultado a 1 de mayo de 2022, disponible en: <https://soundcloud.com/datajournalism/episode-42-conversation-with-brant-houston-stephen-doig>

Koeze, E., Lu, D. y Smart, C. (27 de enero de 2022). Can you gerrymander your party to power? The New York Times. Consultado a 23 de febrero de 2022, disponible en: <https://www.nytimes.com/interactive/2022/01/27/us/politics/congressional-gerrymandering-redistricting-game-2022.html>

Kotenidis, E. y Veglis, A. (2021). Algorithmic Journalism—Current Applications and Future Perspectives, *Journalism and Media.*, 2021; 2(2):244-257. Disponible en: <https://doi.org/10.3390/journalmedia2020014>

Laboratorio de Periodismo (2020). “Guía para introducir la inteligencia artificial en las redacciones”, 08.06.2020 Fundación Luca de Tena, en <https://laboratoriodeperiodismo.org/guia-para-introducir-la-inteligencia-artificial-en-las-redacciones/>

La Nación. (s.f.). Coronavirus: Así Se propaga el virus en el mundo. LA NACION. Consultado a 8 de marzo de 2022, disponible en: <https://www.lanacion.com.ar/el-mundo/coronavirus-asi-se-propaga-virus-mundo-nid2351138/>

Larson, J. y Angwin, J. (23 de mayo de 2016). Machine bias. ProPublica. Consultado a 13 de febrero de 2019, disponible en: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

Lecompte, C. (19 de octubre de 2020). Automation in the newsroom. Nieman Reports. Consultado a 4 de marzo de 2022, disponible en: <https://niemanreports.org/articles/automation-in-the-newsroom/>

Lentacker, A. (2021). The Opioid Documents: A Report on the Politics of the Public Record, *The Social History of Alcohol and Drugs* 2021 35:1, 137-165. Disponible en: <https://www.journals.uchicago.edu/doi/full/10.1086/713409>

Los Angeles Times. (s.f.). California fire map: Active wildfires. Los Angeles Times. Consultado a 2 de febrero de 2022, disponible en: <https://www.latimes.com/wildfires-map/>

Los Angeles Times. (s.f.). Quakebot. Los Angeles Times. Consultado a 6 de enero de 2022, disponible en: <https://www.latimes.com/people/quakebot>

Maldita.es. (Junio de 2021). Disinformation on Whatsapp: maldita.es chatbot and the frequently forwarded attribute, Maldita. Consultado a 15 de Enero de 2022, disponible en:

Marconi, F. (13 de Octubre de 2019). How the Wall Street Journal is using Deep Learning to inform content strategy. Medium. Consultado a 13 de enero de 2022, disponible en: <https://medium.com/the-wall-street-journal/how-the-wall-street-journal-is-using-deep-learning-to-inform-content-strategy-4b4a07090110>

Marconi, F. (2020). *Newsmakers: Artificial Intelligence and the Future of Journalism*. United States: Columbia University Press.

Marconi, F. y Siegman, A. (2017). A day in the life of a journalist in 2027: Reporting meets AI. *Columbia Journalism Review*.

Marconi, F. y Siegman, A. (2017). *The Future of Augmented Journalism: A guide for newsrooms in the age of smart machines*. Associated Press. Disponible en: [https://www.ap.org/assets/files/2017\\_ai\\_guide.pdf](https://www.ap.org/assets/files/2017_ai_guide.pdf)

McGregor, S. (2013). Car hits the mainstream. *Columbia Journalism Review*. Consultado a 16 de enero de 2022, disponible en: [https://archives.cjr.org/data\\_points/computer\\_assisted\\_reporting.php](https://archives.cjr.org/data_points/computer_assisted_reporting.php)

Meyer, P. (1979). Precision Journalism: A Reporter's Introduction to Social Science Methods. United Kingdom: Indiana University Press.

Michel, T. y Rebma. (13 de mayo de 2021). Analysis & Updates: Dating data: How we used multiple dating schemes to provide the most complete picture of the pandemic.

Red Ética (2020). "Coronavirus: 10 virtudes del periodismo responsable ante la pandemia", en Fundación Gabo, Cartagena de Indias-Colombia 17.03.2020 en <https://fundaciongabo.org/es/etica-periodistica/recursos/coronavirus-10-virtudes-del-periodismo-responsable-ante-la-pandemia>

The COVID Tracking Project. Consultado a 10 de febrero de 2022, disponible en: <https://covidtracking.com/analysis-updates/how-we-used-multiple-dating-schemes-to-provide-the-most-complete-picture-of-the-pandemic>

Klasa, A., Wisniewska, A., Stabe, M., Kao, J. S., Smith, A., Blood, D., Madsen, W., Harlow, M., Elliott, O., Burn-Murdoch, J. y Tilford, C. (s.f.). Coronavirus tracker: The latest figures as countries fight the COVID-19 resurgence: Free to read. Financial Times. Consultado a 8 de marzo de 2022, disponible en: <https://www.ft.com/content/a2901ce8-5eb7-4633-b89c-cbdf5b386938>

Mullin, B. (21 de mayo de 2016). Bloomberg EIC: Automation is 'crucial to the future of journalism'. Poynter. Consultado a 2 de febrero de 2019, disponible en: <https://www.poynter.org/tech-tools/2016/bloomberg-eic-automation-is-crucial-to-the-future-of-journalism/>

Newtral. (9 de marzo de 2022). Sanidad dejará de publicar los datos sobre evolución del covid a diario. Newtral. Consultado a 12 de marzo de 2022, disponible en: [m https://www.newtral.es/sanidad-datos-covid-informe/20220305/](https://www.newtral.es/sanidad-datos-covid-informe/20220305/)

Nyc.gov. (16 de mayo de 2018). Mayor de Blasio Announces First-In-Nation Task Force To Examine Automated Decision Systems Used By The City, The Official Website of the City of New York, NYC. Disponible en: <https://www1.nyc.gov/office-of-the-mayor/news/251-18/mayor-de-blasio-first-in-nation-task-force-examine-automated-decision-systems-used-by>

O'Neil, C. (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. United Kingdom: Crown.

OdiséIA. (18 de noviembre de 2020). Automating society report 2020. Algoritmos en el poder, opacidad frente a utilidad social. Medium. Consultado a 24 de enero de 2021, disponible en: <https://medium.com/odiseia/automating-society-report-2020-algoritmos-en-el-poder-opacidad-frente-a-utilidad-social-d9e78319cb2d>

Olaya, V. (2014). Sistemas e Información geográfica. Disponible en:

Pérez-Montoro, M. (2022). Comunicación visual de la información: qué y cómo podemos narrar con datos, Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT). Disponible en: [http://diposit.ub.edu/dspace/bitstream/2445/183791/4/P%C3%A9rez-Montoro%20%282022%29%20Comunicaci%C3%B3n%20visual%20de%20la%20informaci%C3%B3n\\_def.pdf](http://diposit.ub.edu/dspace/bitstream/2445/183791/4/P%C3%A9rez-Montoro%20%282022%29%20Comunicaci%C3%B3n%20visual%20de%20la%20informaci%C3%B3n_def.pdf)

Pitt, F. (2014). Sensors and Journalism Tow Center for Digital Journalism, A Tow/Knight Report New York, NY, p. 15.

Polaino, R. C., Cárdbaba, M., Cirujano, E. y Cambra, U. (2021). Las vacunas contra la Covid en Twitter: redes, mensajes y bots. In Digital media: el papel de las redes sociales en el ecosistema educomunicativo en tiempos de Covid-19. McGraw-Hill Interamericana de España, 487-498.

Prasad, S. (7 de Julio de 2016). This summer, WNYC will investigate New York Heat with sensors and volunteers. Poynter. Consultado a 6 de junio de 2020, disponible en: <https://www.poynter.org/tech-tools/2016/this-summer-wnyc-will-investigate-new-york-heat-with-sensors-and-volunteers/>

Rich, S., Sánchez Díez, M. y Vongkiatkajorn, K. (3 de septiembre de 2019). How to download and use the DEA pain pills database. The Washington Post. Consultado a 5 de diciembre de 2020, disponible en: <https://www.washingtonpost.com/national/2019/07/18/how-download-use-dea-pain-pills-database/>

Rojas, J. L. y Toural, C. (2019). Automated sports journalism. The AnaFut case study, the bot developed by El Confidencial for writing football match reports. Doxa Comunicación, (29), 235-254. Disponible en: [https://repositorioinstitucional.ceu.es/bitstream/10637/10745/2/en\\_m4\\_stamped.pdf](https://repositorioinstitucional.ceu.es/bitstream/10637/10745/2/en_m4_stamped.pdf)

Schmitz Weiss, A. (2016). Sensor journalism: Pitfalls and possibilities. Palabra Clave, 19(4), 1048-1071. Disponible en: [http://www.scielo.org.co/scielo.php?script=sci\\_arttext&pid=S0122-82852016000401048#B14](http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0122-82852016000401048#B14)

Segnini, G. (2008). Introducción al Periodismo Investigativo y al Periodismo Asistido por Computadora, en Agencia de los Estados Unidos para el Desarrollo Internacional (USAID), Guatemala.

Shedden, D. (2 de mayo de 2013). New media timeline (1980). Poynter. Consultado a 6 de febrero de 2021, disponible en: <https://www.poynter.org/archive/2004/new-media-timeline-1980/>

#:~:text=The%20first%20newspaper%20to%20go,The%20Middlesex%20News%2C%20the%20St.

Shedden, D. (24 de noviembre de 2014). Today in media history: CompuServe and the First Online Newspapers. Poynter. Consultado a 6 de febrero de 2021, disponible en: <https://www.poynter.org/reporting-editing/2014/today-in-media-history-compuserve-and-the-first-online-newspapers/>

Shedden, D. (6 de noviembre de 2014). Today in media history: In 1952, a computer helped CBS predict the winner of the presidential election. Poynter. Consultado a 5 de febrero de 2022, disponible en: <https://www.poynter.org/reporting-editing/2014/today-in-media-history-in-1952-a-univac-computer-helped-cbs-news-predict-the-winner-of-the-presidential-election/>

South China Morning Post. (s.f.). Coronavirus: The disease covid-19 explained. South China Morning Post. Consultado a 8 de marzo de 2022, disponible en: <https://multimedia.scmp.com/infographics/news/china/article/3047038/wuhan-virus/index.html>

Stevens, H. (14 de Marzo de 202). These simulations show how to flatten the coronavirus growth curve. The Washington Post. Consultado a 25 de febrero de 2022, disponible en: <https://www.washingtonpost.com/graphics/2020/world/corona-simulator/>

Stevens, H. (14 de Marzo de 2020). Por qué brotes como el del coronavirus crecen exponencialmente y cómo ‘aplanar la curva. The Washington Post. Consultado a 25 de febrero de 2022, disponible en: <https://www.washingtonpost.com/graphics/2020/world/corona-simulator-spanish/>

Stiles, M. (1 de marzo de 2020). Track donations in the L.A. county supervisor's race for the 2nd district. Los Angeles Times. Consultado a 12 de enero de 2022, disponible en: <https://www.latimes.com/projects/county-supervisor-2nd-district-campaign-donations/>

Straits Times. (s.f.). Coronavirus: Countries and regions impacted so far. The Straits Times. Consultado a 8 de marzo de 2022, disponible en: <https://www.straitstimes.com/multimedia/graphics/2020/02/coronavirus-global-numbers/index.html>

Stray, J. (2016). The age of the Cyborg. Columbia Journalism Review. Consultado a 7 de abril de 2021, disponible en: [https://www.cjr.org/analysis/cyborg\\_virtual\\_reality\\_reuters\\_tracer.php](https://www.cjr.org/analysis/cyborg_virtual_reality_reuters_tracer.php)

Swenson, Kyle. (29 de abril de 2020). Millions track the pandemic on Johns Hopkins's dashboard. Those who built it say some miss the real story, The Washington Post. Consultado a 17 de enero de 2022, disponible en: [https://www.washingtonpost.com/local/johns-hopkins-tracker/2020/06/29/daea7eea-a03f-11ea-9590-1858a893bd59\\_story.html](https://www.washingtonpost.com/local/johns-hopkins-tracker/2020/06/29/daea7eea-a03f-11ea-9590-1858a893bd59_story.html)

Süddeutsche Zeitung. (s.f.). Alle daten zum coronavirus. Süddeutsche.de. Consultado a 8 de marzo de 2022, disponible en: <https://www.sueddeutsche.de/projekte/artikel/wissen/coronavirus-die-wichtigsten-daten-in-grafiken-e536967/>

Tagliaferri, L., Tagliaferri, L. (2018). How To Code in Python 3. United States: DigitalOcean.

Tejedor, L. y Dader, José Luis. (2011). El patrimonio del César ante el Perro Guardián: un análisis de Periodismo de Precisión sobre la transparencia económica de los miembros del Gobierno en España. En Estudios sobre el Mensaje Periodístico, volumen 17, número 2, pp. 431-455. Disponible en: <http://revistas.ucm.es/index.php/ESMP/article/view/38124/36876>

The Atlanta Journal-Constitution. (2016). About the AJC's investigation of doctor misconduct. Doctors y Sex Abuse. The Atlanta-Journal Constitution. Consultado a 21 de noviembre de 2022, disponible en: [https://doctors.ajc.com/about\\_this\\_investigation/](https://doctors.ajc.com/about_this_investigation/)

The Atlanta Journal-Constitution. (s.f.). Doctors & Sex Abuse. The Atlanta Journal-Constitution. Consultado a 21 de noviembre de 2022, disponible en: <https://doctors.ajc.com/>

The COVID Tracking Project. (s.f.). About Us. Consultado a 10 de febrero de 2022, disponible en: <https://covidtracking.com/about>

The COVID Tracking Project. (s.f.). State screenshots. The COVID Tracking Project. Consultado a 4 de febrero de 2022, disponible en: <https://screenshots.covidtracking.com/>

The Guardian. (s.f.). Covid World Map: Which countries have the most coronavirus vaccinations, cases and deaths? The Guardian. Consultado a 8 de marzo de 2022, disponible en: <https://www.theguardian.com/world/2021/dec/01/covid-world-map-which-countries-have-the-most-coronavirus-vaccinations-cases-and-deaths>

The Markup (16 de octubre de 2020). The citizen browser project-auditing the algorithms of disinformation. The Markup. Consultado a 14 de enero de 2022, disponible en: <https://themarkup.org/citizen-browser>

The Markup. (s.f.). About Us – The markup. The Markup. Consultado a 20 de marzo de 2021, disponible en: <https://themarkup.org/about>

The Markup. (s.f.). Privacy policy. The Markup. Consultado a 12 de enero de 2022, disponible en: <https://themarkup.org/privacy>

The Marshall Project. (s.f.). The next to die. The Marshall Project. Consultado a 11 de enero de 2022, disponible en: <https://www.themarshallproject.org/next-to-die>

The nerd blog. (s.f.). ProPublica. Consultado a 1 de mayo de 2021, disponible en: <https://www.propublica.org/nerds>

The New York Times. (2020, November 3). Election needles: President. The New York Times. Disponible en: <https://www.nytimes.com/interactive/2020/11/03/us/elections/forecast-president.html>

The New York Times. (s.f.). Coronavirus world map: Tracking the global outbreak. The New York Times. Consultado a 8 de marzo de 2022, disponible en: <https://www.nytimes.com/interactive/2021/world/covid-cases.html>

The New York Times. (22 de noviembre de 2021). The Trouble With the Case Curve During the Holidays. The New York Times. Consultado a 10 de febrero de 2022, disponible en: <https://www.nytimes.com/interactive/2021/11/22/us/covid-data-holiday-averages.html>

The New York Times. (24 de noviembre de 2020). Track coronavirus cases in places important to you. The New York Times. Consultado a 11 de febrero de 2022, disponible en: <https://www.nytimes.com/interactive/2021/us/covid-cases-deaths-tracker.html>

The New York Times. (25 de febrero de 2021). Tracking coronavirus cases at U.S. colleges and Universities. The New York Times. Consultado a 11 de febrero de 2022, disponible en: <https://www.nytimes.com/interactive/2021/us/college-covid-tracker.html>

The New York Times. (27 de junio de 2020). Nearly one-third of U.S. coronavirus deaths are linked to nursing homes. The New York Times. Consultado a 11 de febrero de 2022, disponible en: <https://www.nytimes.com/interactive/2020/us/coronavirus-nursing-homes.html>

The New York Times. (3 de noviembre de 2020). Election needles: President. The New York Times. Consultado a 10 de septiembre de 2021, disponible en: <https://www.nytimes.com/interactive/2020/11/03/us/elections/forecast-president.html>

The New York Times. (s.f.). NYT-COVID-19-bot - overview. GitHub. Consultado a 15 de febrero de 2022, disponible en: <https://github.com/nyt-covid-19-bot>

The New York Times. (s.f.). Nytimes/COVID-19-data: An ongoing repository of data on coronavirus cases and deaths in the U.S. GitHub. Consultado a 15 de febrero de 2022, disponible en: <https://github.com/nytimes/covid-19-data>

The Visual and Data Journalism Team. (s.f.). Covid map: Coronavirus cases, deaths, vaccinations by country. BBC News. Consultado a 8 de marzo de 2022, disponible en: <https://www.bbc.com/news/world-51235105>

The Washington Post. (s.f.). Mapping the worldwide spread of the coronavirus. The Washington Post. Consultado a 8 de marzo de 2022, disponible en: <https://www.washingtonpost.com/graphics/2020/world/mapping-spread-new-coronavirus/>

The Washington Post. (17 de noviembre de 2020). High School Football this week: James Madison at South Lakes. The Washington Post. Consultado a 5 de abril de 2021, disponible en: <https://www.washingtonpost.com/allmetsports/2017-fall/games/football/91619/>

The Washington Post. (20 de enero de 2021). Tracking all of president Trump's false or misleading claims. The Washington Post. Consultado a 12 de enero de 2022, disponible en: <https://www.washingtonpost.com/graphics/politics/trump-claims-database/>

The Washington Post. (20 de Julio de 2019). Follow the Post's investigation of the opioid epidemic. The Washington Post. Consultado a 5 de julio de 2022, disponible en: <https://www.washingtonpost.com/national/2019/07/20/opioid-files/>

The Washington Post. (21 de Julio de 2019). Drilling into the DEA's Pain Pill Database. The Washington Post. Consultado a 4 de diciembre de 2020, disponible en: <https://www.washingtonpost.com/graphics/2019/investigations/dea-pain-pill-database/>

The Washington Post. (29 de abril de 2020 ). Millions track the pandemic on Johns Hopkins's dashboard. Those who built it say some miss the real story, The Washington Post. Consultado a 17 de enero de 2022, disponible en: [https://www.washingtonpost.com/local/johns-hopkins-tracker/2020/06/29/daea7eea-a03f-11ea-9590-1858a893bd59\\_story.html](https://www.washingtonpost.com/local/johns-hopkins-tracker/2020/06/29/daea7eea-a03f-11ea-9590-1858a893bd59_story.html)

Tracy, M. (1 de diciembre de 2019). These reporters rely on public data, rather than secret sources. The New York Times. Consultado a 12 de septiembre de 2021, disponible en: <https://www.nytimes.com/2019/12/01/business/media/open-source-journalism-bellingcat.html>

Tracy, M. (6 de agosto de 2019). The Markup, a tech news site, reinstalls its fired editor as part of a fresh start. The New York Times. Consultado a 20 de marzo de 2021, disponible en: <https://www.nytimes.com/2019/08/06/business/media/the-markup-julia-angwin.html>

Vildrim, P. (9 de abril de 2019). Will ai save journalism -- or kill it? Knowledge at Wharton. Consultado a 5 de mayo de 2020, disponible en: <https://knowledge.wharton.upenn.edu/article/ai-in-journalism/>

VanderPlas, J. (2016). Python Data Science Handbook: Essential Tools for Working with Data. United States: O'Reilly Media.

VV.AA. (2019). News Automation The rewards, risks and realities of 'machine journalism'. WAN-IFRA. Disponible en: [http://immersivaautomation.com/wp-content/uploads/2019/06/WAN-IFRA\\_News\\_Automation-FINAL.pdf](http://immersivaautomation.com/wp-content/uploads/2019/06/WAN-IFRA_News_Automation-FINAL.pdf)

Waller, A. y Lecher, C. (21 de enero de 2022). Help us investigate Facebook Pixel Tracking – the markup. Help Us Investigate Facebook Pixel Tracking – The Markup. Consultado a 14 de enero de 2022, disponible en: <https://themarkup.org/pixel-hunt/2022/01/21/help-us-investigate-facebook-pixel-tracking>

Wang, S. (18 de mayo de 2016). The Hindustan Times is working to build the definitive online source of real-time air quality in all of India. Nieman Lab. Consultado a 6 de junio de 2020, disponible en: <http://www.niemanlab.org/2016/05/the-hindustan-times-is-working-to-build-the-definitive-online-source-of-real-time-air-quality-in-all-of-india/>

WashPostPR. (16 de diciembre de 2021). Arc XP in 2021. The Washington Post. Consultado a 12 de enero de 2022, disponible en: <https://www.washingtonpost.com/pr/2021/12/16/arc-xp-2021/>

Wayback Machine. (s.f.) Wayback Machine. Consultado a 4 de febrero de 2022, disponible en: <https://web.archive.org/>

Willis, D., & Syed, M. (7 de Julio de 2020). Tracking PPP: Search every company approved for Federal Loans. ProPublica. Consultado a 12 de enero de 2022, disponible en: <https://projects.propublica.org/coronavirus/bailouts/>

Wu, J., Watkins, D., Williams, J., Bhagat, S. V., Kumar, H., Gettleman, J., Taylor, R., Davis, L. y Singh, K. D. (17 de diciembre de 2020). Who gets to breathe clean air in New Delhi? The New York Times. Consultado a 10 de abril de 2022, disponible en: <https://www.nytimes.com/interactive/2020/12/17/world/asia/india-pollution-inequality.html>

Zhang, J. (24 de Julio de 2015). Introducing Censusamericans, a Twitter bot for America. FiveThirtyEight. Consultado a 11 de enero de 2022, disponible en: <https://fivethirtyeight.com/features/introducing-censusamericans-a-twitter-bot-for-america/>



# Apéndices



## **Relación de siglas, acrónimos y abreviaturas**

ADM: Procesos de toma de decisiones automatizados.

AP: Associated Press.

API: Application programming interface.

Bot: Robot

ARCOS: Automation of Reports and Consolidated Orders System

CAR: Periodismo asistido por ordenador.

CBS: Columbia Broadcasting System.

CSSE: The Center for Systems Science and Engineering de John Hopkins University

DEA: Agencia Federal Antinarcoóticos

IRE: Investigative Reporters and Editors

OMS: Organización Mundial de las Salud

NICAR: National Institute for Computer-Assisted Reporting

NLP: Natural Language Processing

SARS: Síndrome respiratorio agudo grave

VCS: Control de versiones.

W3C: Wide Web Consortium

WSJ R&D: The Wall Street Journal



## **Glosario**

**Algoritmo:** Conjunto ordenado de operaciones o fórmulas sistemáticas que permite hacer un cálculo para hallar la solución de un problema.

**Bot:** Un bot es una aplicación de software programada por un profesional de la información que ejecuta tareas estructuradas y repetitivas de forma automatizada.

**Control de versiones:** Sistema que guarda los cambios que se producen en un archivo o archivos, creando un historial. El control de versiones permite la colaboración y la revisión del código informático por parte de diferentes profesionales.

**GitHub:** Plataforma o servicio alojado en la nube para compartir y editar código que contiene un sistema de control de versiones denominado Git.

**Git** es un sistema que permite a los profesionales colaborar en un mismo proyecto, editar el código y mantener un historial detallado de todos los cambios que se han producido durante el proceso.

**Open Source:** Se refiere a los programas informáticos cuyo código de programación esta disponible en la web para que cualquier usuario pueda utilizarlo, editarlo o mejorarlo.

**News Bot:** programa informático cuyas tareas estructuradas y repetitivas están relacionadas con la información y la comunicación de noticias de última hora.

Periodismo asistido por ordenador: Popularmente conocido como CAR, por sus siglas en inglés. Disciplina que se sirve de los ordenadores para recopilar y analizar datos que serán la base de historias periodísticas.

Periodismo computacional: Según Diakopoulos, es la disciplina que aplica las técnicas y razonamientos de la computación a actividades periodísticas como la recolección de información, la estructuración de esos datos, así como la comunicación, presentación y diseminación de esa información. La precisión y la veracidad, propias de cualquier pieza periodística, son también son dos principios fundamentales del periodismo computacional.

Periodismo cuantitativo: Disciplina periodística que utiliza las técnicas del análisis cuantitativo, como la matemática y la estadística, para explicar historias o pronosticar eventos futuros como el resultado de unas elecciones. Algunos de los periodistas de *FiveThirtyEight* o *The Markup* se definen así mismo como periodistas cuantitativos.

Periodismo de algoritmos: Definido por Diakopoulos como *algorithmic accountability reporting*. Consiste en la cobertura informativa de los algoritmos con el objetivo de rendir cuentas sobre sus efectos en la sociedad. La cobertura informativa de los algoritmos busca identificar los algoritmos en los que la sociedad no repara pero que rigen o influyen en el día a día de sus vidas. El periodista, por tanto, investiga y estudia las características e impacto de estos algoritmos en las personas. El objetivo de estas coberturas o investigaciones es la rendición de cuentas de los actores públicos o privados que han diseñado el algoritmo o han promovido su implantación.

Periodismo de precisión: Disciplina que defiende la aplicación de métodos de otras ciencias para contar historias periodísticas. El periodista y profesor de periodismo, Philip Meyer, profundiza en esta especialidad en su libro “Precision Journalism: Reporter’s Introduction to Social Sciences.”. Meyer propuso y desarrolló el análisis de datos para contar historias periodísticas. Combinó sus conocimientos periodísticos con técnicas provenientes de otras disciplinas como la sociología, la estadística o la ingeniería informática.

Periodismo de sensores: Esta disciplina periodística utiliza sensores para generar, recopilar e interpretar datos con el fin de resolver hipótesis periodísticas y contar historias basadas en esos datos. El profesional de la información puede analizar datos de sensores de terceros, crear y diseñar sus propios sensores o servirse de la ayuda de los lectores para desplegar los sensores que servirán para recopilar los datos que formarán parte de la historia.

Repositorio: Archivo o carpeta de archivos de código de programación almacenados en la nube que los usuarios utilizan para colaborar, trabajar simultáneamente y compartir código.

Software libre: Conjunto de programas informáticos que se rigen por una licencia libre. Es decir, los usuarios tienen la libertad de ejecutar, copiar, distribuir, estudiar, modificar y mejorar los programas de software libre.

Scraping: Técnica para extraer información o datos de una web de forma automatizada y exportar esos datos de forma estructurada.

## **Índice de tablas**

Tabla 1. Nuevas tareas del periodista como consecuencia de la absorción de diferentes procesos tecnológicos.

Tabla 2. Lenguajes de programación más comunes para la visualización de datos en los medios de comunicación.

Tabla 3. Aplicaciones de algunas de las herramientas más populares para la creación de visualizaciones de datos en el sector del periodismo.

## Índice de figuras

Figura 1. Artículo publicado el 15 de Octubre de 1952 en el diario Philadelphia Evening Bulletin sobre el uso del ordenador QUES Remington Rand UNIVAC por la cadena CBS para predecir los resultados de la elección presidencial de Estados Unidos de 1952.

Figura 2. Investigación Freeloading Insurors Get your \$59,000 del periodista Philip Meyer publicada en la portada del diario estadounidense *The Miami Herald* el 27 de Septiembre de 1959.

Figura 3. Resultados de la encuesta de Data Visualization Society a la pregunta sobre qué tecnología o lenguaje de programación es el más usado por los profesionales de la información.

Figura 4. Agujas diseñadas y utilizadas por el diario estadounidense *The New York Times* para ofrecer contexto y predecir los resultados de las elecciones presidenciales estadounidenses de 2020.

Figura 5. Apuesta visual de *The Washington Post* para ofrecer contexto y mostrar los diferentes escenarios posibles durante el recuento de votos en una elección.

Figura 6. Clasificación propuesta para el análisis de bots en el ámbito del periodismo.

Figura 7. Reproducción de una conversación con el lector del bot conversacional y asistente personal de finanzas desarrollado por *The Washington Post*.

Figura 8. Resultado de la búsqueda de la web de la Universidad Complutense de Madrid (ucm.es) en la herramienta Blacklight.

Figura 9. Resultado de la búsqueda del portal de datos de Nevada sobre la COVID-19 en la herramienta *Blacklight*.

Figura 10. Resultado de la búsqueda de la web de The Markup (themarkup.org) en la herramienta Blacklight

Figura 11. Mapa del número de opioides distribuidos por año en cada condado de Estados Unidos.

Figura 12. Extracto del artículo *A Plastic Bag's 2,000-Mile Journey Shows the Messy Truth About Recycling* publicado por *Bloomberg* que muestra el recorrido de las tres bolsas de plástico rastreadas mediante el uso de sensores diminutos.

Figura 13. Diagrama del mapa de metro de Londres publicado por *Financial Times* con los niveles de contaminación por tramos de vía.

Figura 14. Introducción del artículo *Who Gets to Breathe Clean Air in New Delhi?* publicado por *The New York Times*.

Figura 15. Proceso de subida de datos a la herramienta de creación de gráficos Datawrapper.

Figura 16. Proceso de chequeo y edición de datos de la herramienta de creación de gráficos Datawrapper.

Figura 17. Proceso de selección de gráficas y visualizaciones de datos de la herramienta de creación de gráficos Datawrapper.

Figura 18. Aspecto del tracker de *South China Morning Post* de impacto de la COVID-9 el 29 de enero de 2020.

Figura 19. Aspecto del tracker de *South China Morning Post* de impacto de la COVID-9 el 30 de enero de 2021.

Figura 20. Tipos de gráficos y mapas utilizados en los diferentes trackers estudiados en esta disertación.

Figura 21. Esquema de tareas manuales y automáticas realizadas por los periodistas del diario *The New York Times* para el procesamiento, edición y publicación de datos de la COVID-19.

Figura 22. Visualización de datos con la simulación de personas enfermas, sanas y recuperadas del artículo de *The Washington Post* titulado *Why outbreaks like coronavirus spread exponentially, and how to “flatten the curve”*.

Figura 23. Artículo *Can you Gerrymander your party to power?* publicado por *The New York Times*.

Figura 24. Escena del juego *The Uber Game* desarrollado por *Financial Times*.



# **Anexos**



## **Anexo I. Formulario para entrevistas a profesionales de la información sobre su trabajo de análisis y la visualización de datos de la COVID-19.**

A continuación, se reproduce el formulario utilizado para recabar datos sobre la cobertura visual de la COVID-19 a nivel mundial utilizado en el estudio *Effectiveness, clarity and impact of visual journalism in the time of a pandemic* <sup>212</sup>. El formulario online está disponible en la siguiente dirección web: <https://forms.gle/MWQi2ycDBipam2YW8>

Este estudio académico fue posible gracias a una beca 2020 Knight Fellow <sup>213</sup> del centro de estudios Tow Center for Digital Journalism de la Universidad de Columbia. Para la elaboración del estudio he trabajado con mi compañero y doctorando en la facultad Columbia Journalism School de la Universidad. Columbia, Javier Sauras.

El objetivo del estudio del estudio era analizar la claridad, efectividad e impacto de las visualizaciones de datos y trackers publicados en los principales medios de comunicación del mundo en relación a la cobertura de la COVID-19. Como parte de esa investigación, se utilizó este formulario para recoger experiencias personales y opiniones de los profesionales de la información con el fin de comprender las diferentes estrategias que se han adoptado para informar sobre la pandemia.

Las respuestas a esta encuesta y las conclusiones del mismo han sido claves para dar forma y completar el CAPÍTULO V: Automatización en periodismo visual.

---

<sup>212</sup> La información y propuesta completa del estudio *Effectiveness, clarity and impact of visual journalism in the time of a pandemic* elaborado para Tow Center for Digital Journalism de la Universidad de Columbia está disponible en: <https://towcenter.columbia.edu/content/research-projects-0>

<sup>213</sup> Más detalles sobre la beca en la web de Tow Center for Digital Journalism de la Universidad de Columbia: <https://towcenter.columbia.edu/content/adrian-blanco-ramos>

## ***Effectiveness, clarity and impact of visual journalism in the time of a pandemic***

This study aims to analyze the effectiveness, the clarity and the impact of the different visual stories, data visualizations and trackers published on the main U.S. and international media to cover the covid-19 pandemic. Reporting on a complex news event such as a pandemic implies to produce clear, simple and explanatory graphics, so this research will also look to describe the effect that these visual pieces have produced on the audience.

As a consequence of the coronavirus pandemic, and despite their time consuming production, explanatory visual stories flourished in the homepages of main newspapers and publications. Although reporting on these numbers could look as straightforward (most of the reporting revolves around cases, deaths and tests), a few issues surrounding the access, production and interpretation of these numbers arise.

This study will address the different strategies that graphics teams have followed to report on these numbers. The scope of the study will be focused on a selection of graphics and visual stories published around the world depicting original and successful stories but also pitfalls. The research will also analyze the effect on the audience of these different pieces published during the pandemic era. Some of the aspects the researchers will look into are how these graphics helped the audience to get informed about covid-19, if they break any of their presumptions, as well as identifying which feelings evoked in them.

<https://towcenter.columbia.edu/content/research-projects-0>

## **Tell us about you**

All answer to the questions of this section are optional. Please, leave us your e-mail if you do not mind that we get in touch with you again to ask you more questions if needed.

- Role
- Media
- Background and experience

## Journalism and covid-19

Tell us more about the variables you have used to report on the pandemic and the different decisions you have made to choose a graphic, a data point or a measure over another.

- Before you started reporting on COVID-19, how were you following the evolution of the pandemic? What media outlets were you reading? What sources were you using?
- Once you started reporting on COVID-19, what were your main sources for getting data? What's been your inspiration for presenting the data?
- Which are the variables that your media outlet has focused on the most when reporting about the coronavirus?
  - Cases
  - Cases per capita
  - Deaths
  - Deaths per capita
  - Tests
  - Tests per capita
  - Positivity rate
  - Excess deaths
  - Other
- Why?

- What is the time measure your media outlet has usually used to show covid numbers?  
Daily, weekly, monthly, accumulated.

- Daily
- Weekly
- Biweekly
- Monthly
- Accumulated
- Other

- Why?

- When reporting about the pandemic, have you used...

- local level data
- county-level data
- state
- country
- region
- world
- Other

- Why?

- Which is the type of graphics you have used the most to explain the variables above?
  - Multiline charts
  - Line charts in small multiples
  - Bar charts
  - Bar charts in small multiples
  - Stream graphs
  - Maps
  - Cartograms
  - Other:
- Describe why.

## Questions about your media

- Is there any particular variable that your media outlet has associated with a particular graphic over time? (For example: using a line chart to show accumulated cases)
- What are the tracker stories that your media outlet has produced about the coronavirus? Please, add links and describe the main features or purpose behind these trackers...
- What are the most visited graphic stories that your media has produced about the coronavirus?
- What are the featured graphic stories that your media has produced about the coronavirus? (Neither a tracker nor the most visited ones, the stories that your media put so much effort into it because they needed to be told)
- Are there any particular graphics that your media has matched to a particular variable? Why?
- In which cases have you visualized the data over a map?
- Let's focus on the trackers that your media has published over the past months. Can you tell us what was the reasoning behind choosing the particular graphics and variables you are showing?

## **Challenges and personal experience**

- What are the main challenges that you have encountered covering the pandemic? Is there anything particularly related to data and visualizations?
- Could you take us through the main challenges that you encountered as a reporter? Give us examples of successes and pitfalls. During this time, what has worked best, and what has not worked?
- What have you learned and what would you wish you have done better?
- After several months of reporting on COVID-19, what are your main take aways?
- Is there anything else you would like to add?



## **Anexo II. Artículo preliminar del estudio *Effectiveness, clarity and impact of visual journalism in the time of a pandemic*.**

A continuación, se reproducen en sus totalidad las conclusiones preliminares del estudio *Effectiveness, clarity and impact of visual journalism in the time of a pandemic* <sup>214</sup> publicado en el boletín semanal del centro de estudios Tow Center for Digital Journalism de la Universidad de Columbia. El artículo está disponible en su totalidad en la siguiente dirección web: <https://mailchi.mp/95bdcee6e0d7/weekly-updates-on-covid-19s-impact-on-journalism-2083305?e=e8e508e35f>

Este estudio académico fue posible gracias a una beca 2020 Knight Fellow <sup>215</sup> del centro de estudios Tow Center for Digital Journalism de la Universidad de Columbia. Para la elaboración del estudio he trabajado con mi compañero y doctorando en la facultad Columbia Journalism School de la Universidad. Columbia, Javier Sauras.

Las conclusiones de esta publicación han sido claves para dar forma y completar el CAPÍTULO V: Automatización en periodismo visual.

---

<sup>214</sup> La información y propuesta completa del estudio *Effectiveness, clarity and impact of visual journalism in the time of a pandemic* elaborado para Tow Center for Digital Journalism de la Universidad de Columbia está disponible en: <https://towcenter.columbia.edu/content/research-projects-0>

<sup>215</sup> Más detalles sobre la beca en la web de Tow Center for Digital Journalism de la Universidad de Columbia: <https://towcenter.columbia.edu/content/adrian-blanco-ramos>

## **Drawing the curve: Data visualization and COVID-19**

### **A pandemic year in graphics**

By Adrián Blanco and Javier Sauras

When Wuhan’s municipal health commission issued an urgent notice on December 30, 2019 asking local hospitals to report and track cases of “pneumonia of unknown cause,” few newspapers were paying attention. But in Hong Kong, where the news invoked memories of the 1997 avian influenza and the Severe Acute Respiratory Syndrome (SARS) of 2003, health authorities and media professionals reacted quickly. The city’s department of health alerted frontline medical staff and increased temperature screenings at every border checkpoint. Meanwhile, Hong Kong’s press began scanning social media for clues about the disease and reaching out to correspondents in mainland China. Soon, marquee international media companies with a heavy presence in China, such as The New York Times, BBC, The Guardian, Financial Times, and Washington Post followed suit.

As reporters learned more about the behavior and consequences of the virus, the data they tracked and the ways they published them changed. With China being the epicenter of the pandemic, the first data pieces were heavily focused on China. South China Morning Post was one of the first media outlets to explain the pandemic using a visual angle.

Meanwhile, in the U.S., a small team at John Hopkins University was the first to start collecting data and numbers of the global pandemic. They took the lead trying to put the numbers in context and quickly became the go-to source for reporters for data on cases and deaths. But data journalists had to be cautious when using the first iterations of such publicly available data due to errors and reporting inconsistencies between countries. Yet in the first months of the pandemic, almost all data trackers, graphics, and stories were fed by John Hopkins University data. But manual labor and inconsistency issues required extra verification and led to human errors, which made some newsrooms cautious against using the first few iterations of such publicly available data.

By the time the World Health Organization officially declared a global pandemic, the Western press was already publishing data trackers to show cases, deaths, and spread both on a global scale and at the national level. Data trackers flourished in the online front pages of newspapers and media outlets, helping readers understand the pandemic from a visual perspective. In the U.S., The New York Times was the first newsroom to create their own database down to the county level. There were also trackers that collected data not by location but by demographics. The Marshall Project gathered information and statistics on coronavirus cases in prisons to keep facilities accountable. As the multimedia editor of a large international media conglomerate told us, “the COVID-19 pandemic was, perhaps, the single biggest data journalism story of our generation.”

Data visualization journalism is an interdisciplinary practice that combines, at the very least, research, storytelling, statistical analysis, and design skills. Its main purpose is to communicate information clearly and effectively with graphics. And never had data visualization been so meaningful as it became in 2020. For the first time in history, millions of people around the globe tracked line graphs and column charts every day, perused over heat maps, and followed spikes and dips with unprecedented interest. Before the COVID-19 outbreak, concepts like “flattening the curve,” “new wave,” or “R number,” were only privy to a cadre of experts. Now, they have almost become colloquial.

In 2020 we learned that it’s easier to talk about flattening the curve if we have seen it represented before; to grasp the full meaning of the “R number” and the herd immunity threshold if they are paired with graphics. In a year when the evolution of a global virus monopolized the news cycle, many media outlets began to put data visualizations front and center of their daily operations.

To understand the outsized role of data visualizations on media and journalism over the past year, we interviewed over 50 journalists, developers, designers, editors, and scholars who were either covering the pandemic or studying that coverage. We consulted with media professionals working for 40 different organizations in 20 countries spanning six continents.

We questioned our interviewees about the primary data sources they relied upon; their inspiration and influences; the variables, measures, trackers and graphics they used; their most pressing challenges and successes; and the lessons they would apply in the future. Our forthcoming report explores that evolution through the experiences of its protagonists. It also studies the various strategies, graphs, and trackers media professionals followed to report on the pandemic.

## **Key Findings**

- **Location:** Physical proximity to a pandemic hot spot played a crucial role in our interviewees' work. Although they mainly worked from home in 2020, the closer journalists were to the virus, the earlier they reported on it. And the earlier they started reporting on COVID-19, the more original and influential to other media professionals their work was.
- **The rise of local data:** Our interviewees reported that audiences preferred local and national stories over international features. Their news organizations' primary sources were also usually national statistics agencies.
- **Uncertainty:** COVID-19 was an unexpected news event with very little modern precedent for newsroom coverage. With that, one common feeling shared by journalists was the lack of knowledge and understanding of a global pandemic at the beginning of it. Almost overnight, they had to adapt themselves and learn about the virus while they were also covering it.
- **Scarcity and redundancy:** From the beginning, scarcity and centralization of reliable sources caused almost every international, national, and even local media outlet to resort to the same data pool. Publishing the same data from the same sets of information led to redundancy and homogeneity, resulting in a kind of formulaic coverage.

- **Automation:** Those who were able to automate their trackers and visualizations found more time to produce original reporting and pursue investigative stories.
- **Simplicity:** Explaining the pandemic meant explaining something new for everyone. Most journalists interviewed for this report agreed on the importance of visualizing in a simple but effective way, such as with bar charts, weekly average lines, and bubble maps. These also became the most popular types of visualizations.
- **Experts:** Many interviewees underscored the importance of having reliable, steady epidemiologists and statisticians on call before the pandemic. Experts were important in understanding the data and analyzing it, which gave journalists more time and freedom to push back against formulaic reporting.
- **Expressing error:** The gaps, the caveats, and the limited availability of data required reporters to be cautious. Many reported that they made a point to include the possibility of error and variability within their visualization. At the same time, not all media organizations and researchers presented the margin of error, often citing a cost-benefit analysis of providing vital information over perfectionism.
- **Professional Growth:** Many data journalists and visualizers reported feeling more valued today inside and outside their newsrooms than before the pandemic began. They believed their work was more important and better understood in 2020 than years prior.

### **Findings on data sources**

With the exception of those based in East Asia, most of the professionals we interviewed were following the evolution of the coronavirus spread through major US and UK-based media outlets. Johns Hopkins University (JHU), the European Centre for Disease Prevention and Control (ECDC), wire services and Twitter were also preferred sources. The first data many newsrooms used came from the [JHU dashboard](#), a regularly updated tracker that stood

out because “it was well sourced and verifiable with sources on the ground,” said an editor based in Qatar. However, none of these sources was entirely reliable.

Some journalists quickly realized that even data from the most trustworthy sources, such as JHU and the ECDC, had issues with accuracy and timeliness, which limited their use. But many others didn’t recognize these problems or couldn’t find suitable alternatives.

Even as outlets and research organizations began to build their own trackers, hardly anyone approached primary sources like doctors and hospitals to produce original databases. Our interviewees cited time constraints, and lack of skills and resources as reasons why data sourcing was narrow in the early days of the pandemic.

### **Data variables**

Most coronavirus data trackers include cases and deaths variables to show the spread of the virus. These types of visual data trackers became popular early on because they were simple to follow. But across the world, cases (and therefore charts and graphics) were not always accurate because of the variation in availability of tests and vast underreporting in many countries. Deaths caused by the virus were not always precise due to the lag on reporting them and the unnoticed deaths at the beginning of the pandemic.

Trackers that followed cases and death were more popular than those that measured hospitalizations or incidence rates. By nature, this data is more complex to measure and explain, making it harder to represent visually. And despite hospitalizations being one of the best ways to explain the evolution of the pandemic and its risks, this data wasn't widely available until much later in the pandemic.

### **Conclusion**

Covering the pandemic—one of the most unprecedented news events in modern memory—was a tremendous challenge for data professionals all over the world. It was also an

opportunity for data and visual journalism to populate the homepages of nearly every single media outlet across the globe. As the demand and need for data visualizations of the pandemic grew, so too did this type of journalism's influence. To explain the pandemic in accessible ways, reporters developed a very particular visual language summarizing the spread of the virus, while also explaining risks and distilling new terms such as "herd immunity" and "flattening the curve." The recognition and importance that analysis and visualization has received during the pandemic in media organizations should become a powerful argument to tell other data heavy stories in the future.

**Editor's Note: A full version of this report is set to be released through the Tow Center later this summer.**



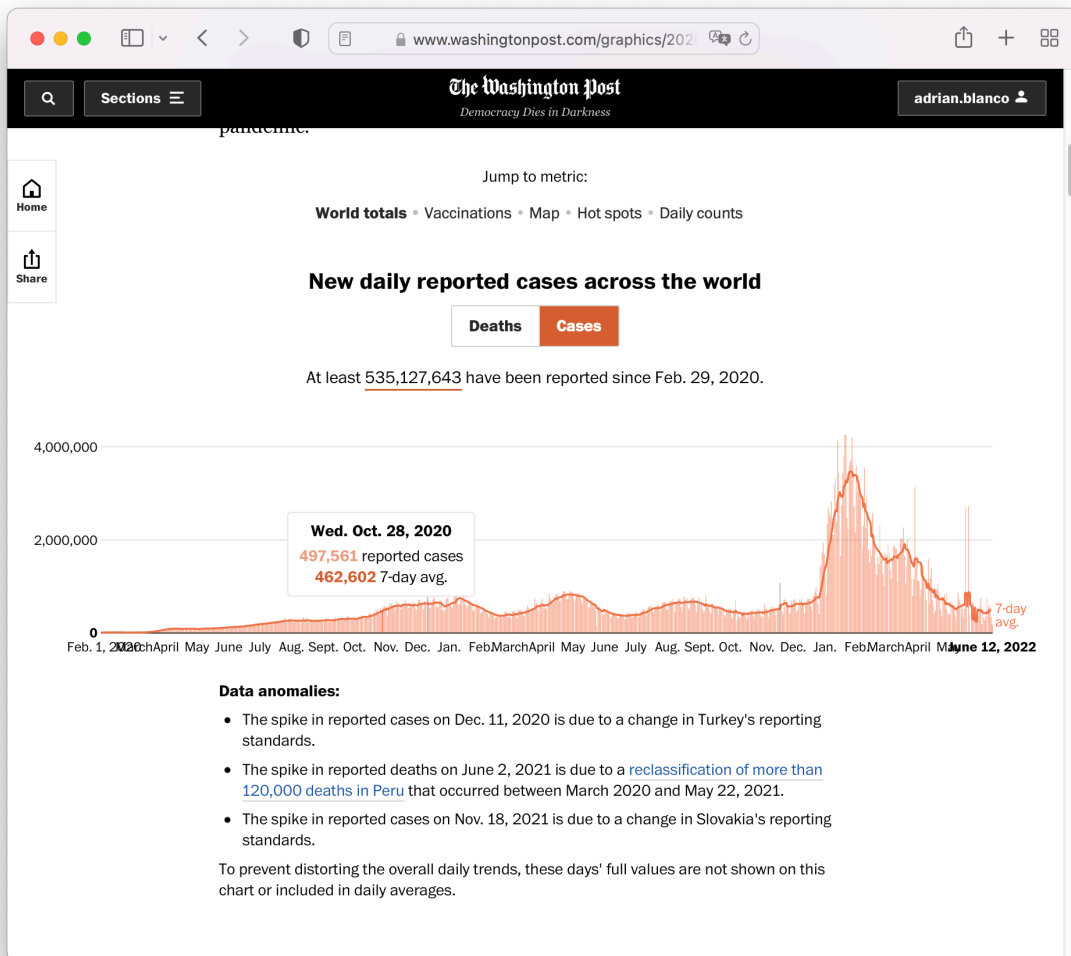
### **Anexo III. Trackers seleccionados para el estudio de la automatización en la cobertura de la COVID-19.**

En las siguientes páginas se reproduce la base de datos elaborada con diferentes trackers o paneles de datos del coronavirus para el estudio y análisis de la cobertura de la pandemia en la automatización de procesos aplicada al periodismo visual. Los criterios de selección de estos trackers son:

- Criterio de automatización: Algunos o todos los procesos de procesamiento, análisis y visualización de datos del tracker han sido automatizados por los profesionales de los datos.
- Criterio de contenidos: Los trackers seleccionados analizan al menos dos variables del virus (casos positivos y fallecidos) y cuentan con al menos tres tipos de visualizaciones, entre gráficos y mapas, en la pieza periodística.
- Criterio de diversidad geográfica: Dentro de lo posible, se ha buscado variedad geográfica a la hora de la selección de los medios con el fin de comparar los trackers desarrollados por medios de comunicación en diferentes continentes.
- Criterio de relevancia y popularidad: Se han seleccionado algunos de los trackers más populares para lectores y expertos de la visualización de datos.
- Criterio de tamaño: se han seleccionado medios de similares características y tamaño. Todos ellos son medios grandes con recursos.

Las conclusiones extraídas de esta base de datos han sido claves para completar el CAPÍTULO V: Automatización en periodismo visual. La consulta y análisis de estos trackers se realizó en marzo de 2022.

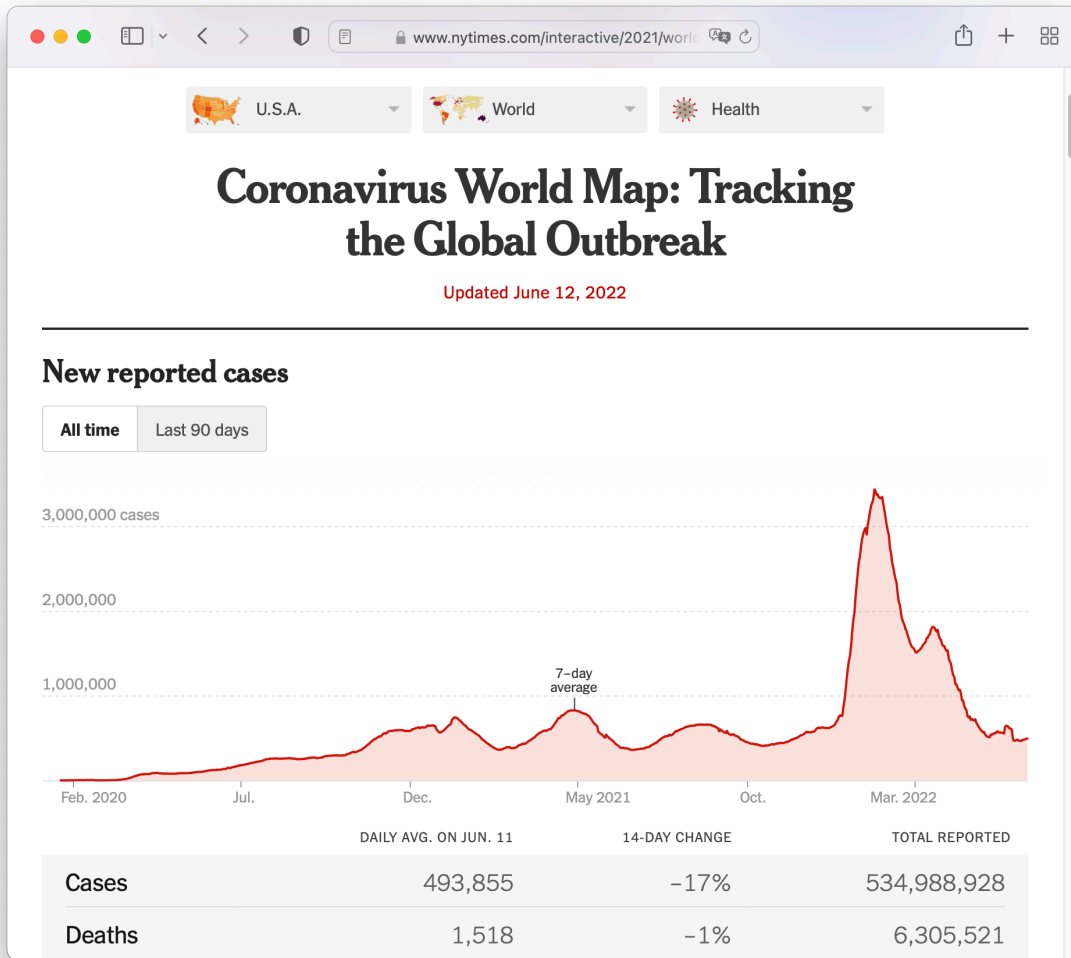
<b>Medio</b>	The Washington Post	<b>País</b>	Estados Unidos
<b>Titular</b>	Tracking covid-19 cases, deaths and vaccines worldwide		
<b>Descripción</b>	More than 6,304,000 people have died from coronavirus worldwide, and more than 535 million cases have been reported.		
<b>Fuentes utilizadas</b>		<b>Tipo de datos</b>	
Data on deaths and cases come from The Washington Post's reporting and the Center for Systems Science and Engineering at Johns Hopkins University. World vaccinations data comes from Our World in Data and The Washington Post's reporting. Country population and economic classification data are from the World Bank. Taiwan population data is from the Taiwan Statistical Bureau.		<ul style="list-style-type: none"> <li>• Casos</li> <li>• Fallecidos</li> <li>• Vacunas</li> </ul>	
<b>Tipos de gráficos</b>		<b>Tipos de mapas</b>	
<ul style="list-style-type: none"> <li>• Barras</li> <li>• Línea (media semanal)</li> <li>• Beeswarm</li> <li>• Tablas</li> </ul>		<ul style="list-style-type: none"> <li>• Burbujas</li> </ul>	
<b>Visualiza incertidumbre</b>		<b>Otros detalles</b>	
Sí, en el titular. Incluye además media semanal para ajustar los datos.		Permite mostrar números totales y ajustados por población. Muestra tanto datos semanales como acumulados.	
<b>Dirección web</b>			
<a href="https://www.washingtonpost.com/graphics/2020/world/mapping-spread-new-coronavirus/">https://www.washingtonpost.com/graphics/2020/world/mapping-spread-new-coronavirus/</a>			



Fuente: *The Washington Post*.

The Washington Post. (s.f.). Mapping the worldwide spread of the coronavirus. The Washington Post. Consultado a 8 de marzo de 2022, disponible en: <https://www.washingtonpost.com/graphics/2020/world/mapping-spread-new-coronavirus/>

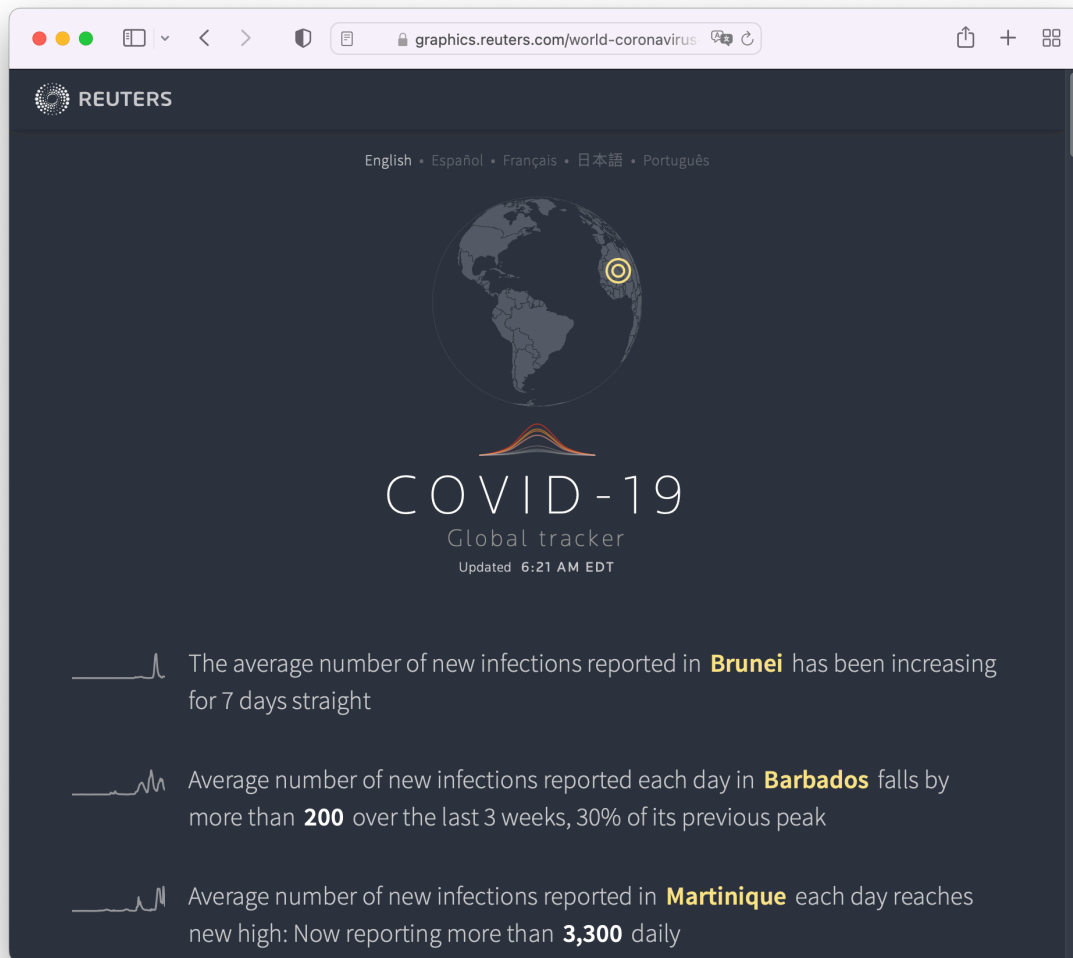
<b>Medio</b>	The New York Times	<b>País</b>	Estados Unidos
<b>Titular</b>	Coronavirus World Map: Tracking the Global Outbreak		
<b>Descripción</b>	Panel de datos con el avance del virus que además de las visualizaciones incluye una serie de actualizaciones en texto.		
<b>Fuentes utilizadas</b>		<b>Tipo de datos</b>	
Data for all countries except the United States comes from the Center for Systems Science and Engineering at Johns Hopkins University. United States data comes from state and local health officials and is collected by The New York Times. Population data from the World Bank and U.S. Census Bureau. Data for some countries, like the United States, Denmark, France and the Netherlands, include counts for overseas territories. The New York Times has found that official tallies in more than thirty countries have undercounted deaths during the coronavirus outbreak because of limited testing availability.		<ul style="list-style-type: none"> <li>• Casos</li> <li>• Fallecidos</li> </ul>	
<b>Tipos de gráficos</b>		<b>Tipos de mapas</b>	
<ul style="list-style-type: none"> <li>• Barras</li> <li>• Línea (media semanal)</li> <li>• Línes (comparativa entre regiones)</li> </ul>		<ul style="list-style-type: none"> <li>• Coropletas</li> </ul>	
<b>Visualiza incertidumbre</b>		<b>Otros detalles</b>	
Utiliza media semanal para ajustar los datos.		Puntualiza "The Times has identified reporting anomalies or methodology changes in the data" y explica las anomalías.	
<b>Dirección web</b>			
<a href="https://www.nytimes.com/interactive/2021/world/covid-cases.html">https://www.nytimes.com/interactive/2021/world/covid-cases.html</a>			



Fuente: *The New York Times*.

The New York Times. (s.f.). Coronavirus world map: Tracking the global outbreak. The New York Times. Consultado a 8 de marzo de 2022, disponible en: <https://www.nytimes.com/interactive/2021/world/covid-cases.html>

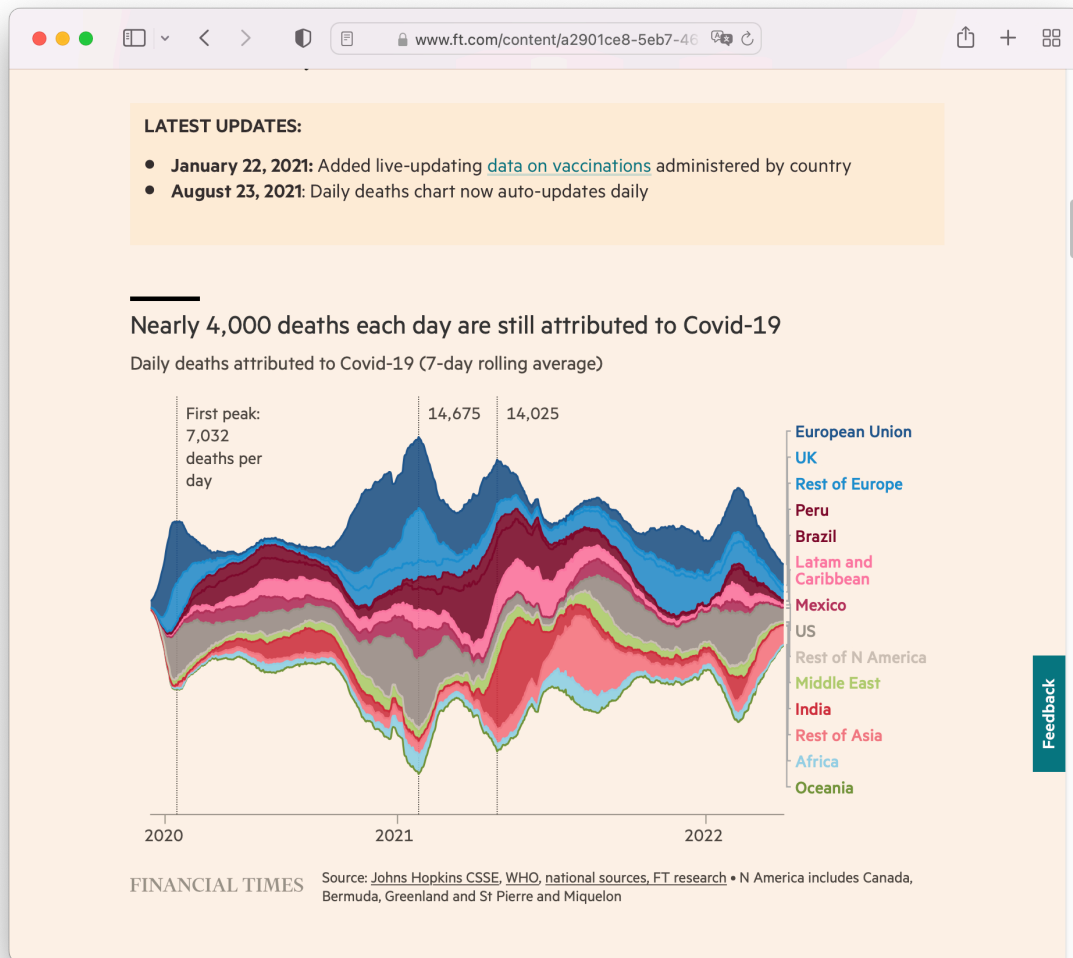
<b>Medio</b>	Reuters	<b>País</b>	Canadá
<b>Titular</b>	COVID-19 Global tracker		
<b>Descripción</b>	Reuters is collecting daily COVID-19 infections and deaths data for 240 countries and territories around the world, updated regularly throughout each day. Every country reports those figures a little differently and, inevitably, misses undiagnosed infections and deaths. With this project we are focusing on the trends within countries as they try to contain the virus' spread, whether they are approaching or past peak infection rates, or if they are seeing a resurgence of infections or deaths.		
<b>Fuentes utilizadas</b>		<b>Tipo de datos</b>	
Local state agencies, local media, Oxford Coronavirus Government Response Tracker, Our World in Data, The World Bank, Reuters research		<ul style="list-style-type: none"> <li>• Casos</li> <li>• Fallecidos</li> <li>• Vacunas</li> </ul>	
<b>Tipos de gráficos</b>		<b>Tipos de mapas</b>	
<ul style="list-style-type: none"> <li>• Barras</li> <li>• Línea (media semanal)</li> <li>• Tablas</li> </ul>		<ul style="list-style-type: none"> <li>• Localizador de país</li> <li>• Spike map</li> </ul>	
<b>Visualiza incertidumbre</b>		<b>Otros detalles</b>	
Utiliza media semanal para ajustar los datos.		Traducido a cinco idiomas Páginas individuales de cada uno de los países con más detalles y comparativas con el rest de países	
<b>Dirección web</b>			
<a href="https://graphics.reuters.com/world-coronavirus-tracker-and-maps/">https://graphics.reuters.com/world-coronavirus-tracker-and-maps/</a>			



Fuente: *Reuters*.

Bhatia, G., Dutta, P. K. y McClure, J. (s.f.). Covid-19: The Latest Global Statistics, charts and maps. Reuters. Consultado a 8 de marzo de 2022, disponible en: <https://graphics.reuters.com/world-coronavirus-tracker-and-maps/>

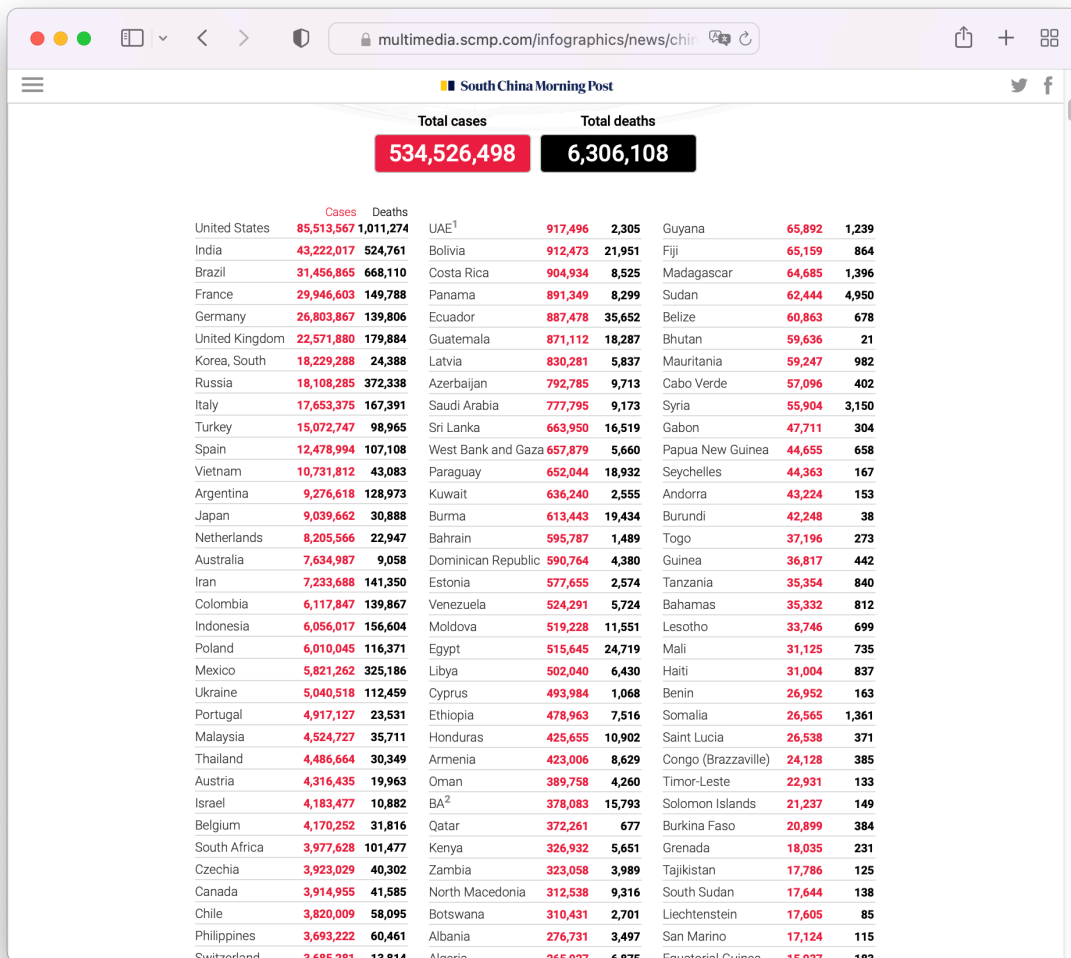
<b>Medio</b>	Financial Times	<b>País</b>	Reino Unido
<b>Titular</b>	Coronavirus tracker: the latest figures as countries fight the Covid-19 resurgence		
<b>Descripción</b>	The FT analyses the scale of outbreaks and tracks the vaccine rollouts around the world		
<b>Fuentes utilizadas</b>		<b>Tipo de datos</b>	
<p>Unless otherwise stated below, the data used for cases and deaths in these charts comes from the Johns Hopkins University Center for Systems Science and Engineering, and reflects the date that cases or deaths were recorded, rather than when they occurred. Data for the US, its individual states, Puerto Rico, Guam, American Samoa, the US Virgin Islands and the Northern Mariana Islands is calculated from county-level data compiled by the Johns Hopkins CSSE. The full excess mortality dataset used for this analysis is freely available for download on Github. It is compiled from data originally produced by official statistics agencies or civil registries in each of the jurisdictions mentioned. The full list of sources is also available on our Github repository. Unless otherwise specified, vaccination data is compiled by Our World in Data, or, where this is the most recent available, the World Health Organization.</p>		<ul style="list-style-type: none"> <li>• Casos</li> <li>• Fallecidos</li> <li>• Exceso de fallecidos</li> <li>• Vacunas</li> </ul>	
<b>Tipos de gráficos</b>		<b>Tipos de mapas</b>	
<ul style="list-style-type: none"> <li>• Streamgraph</li> <li>• Barras</li> <li>• Línea (media semanal)</li> <li>• Líneas</li> <li>• Calendario</li> </ul>		No incluye mapas	
<b>Visualiza incertidumbre</b>		<b>Otros detalles</b>	
<p>Utiliza media semanal para ajustar los datos. Analiza y visualiza exceso de fallecidos para ofrecer una aproximación más exacta al impacto de la pandemia.</p>			
<b>Dirección web</b>			
<a href="https://www.ft.com/content/a2901ce8-5eb7-4633-b89c-cbdf5b386938">https://www.ft.com/content/a2901ce8-5eb7-4633-b89c-cbdf5b386938</a>			



Fuente: *Financial Times*.

Klasa, A., Wisniewska, A., Stabe, M., Kao, J. S., Smith, A., Blood, D., Madsen, W., Harlow, M., Elliott, O., Burn-Murdoch, J. y Tilford, C. (s.f.). Coronavirus tracker: The latest figures as countries fight the COVID-19 resurgence: Free to read. Financial Times. Consultado a 8 de marzo de 2022, disponible en: <https://www.ft.com/content/a2901ce8-5eb7-4633-b89c-cbdf5b386938>

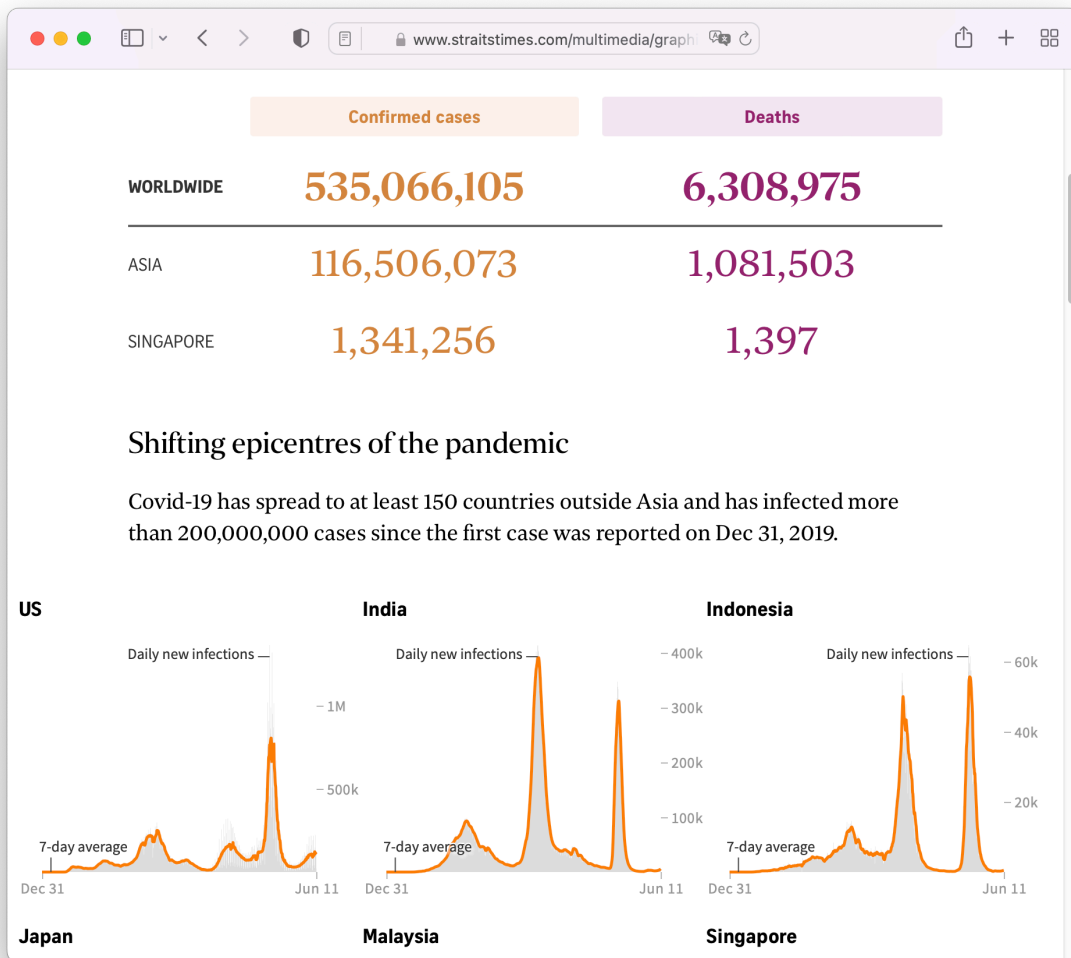
<b>Medio</b>	South China Morning Post	<b>País</b>	China
<b>Titular</b>	Coronavirus: the disease Covid-19 explained		
<b>Descripción</b>	The Covid-19 outbreak caused China to put much of the country into lockdown but the virus has spread rapidly around the world. Many countries are quarantining travellers from overseas. Find out how it all began.		
<b>Fuentes utilizadas</b>		<b>Tipo de datos</b>	
OpenStreetMap, State media, Maps4News, National Health Commission of the PRC; Local Municipal Health Commission; China Health Statistics Yearbook 2018; US' Centres for Disease Control and Prevention (CDC),The Lancet, US Food and Drug Administration (FDA), WHO Summary of probable Sars cases and China's National Health Commission; Baidu Qianxi; BNO News; Flight Master		<ul style="list-style-type: none"> <li>• Casos</li> <li>• Fallecidos</li> <li>• Recuperados</li> </ul>	
<b>Tipos de gráficos</b>		<b>Tipos de mapas</b>	
<ul style="list-style-type: none"> <li>• Gráficos de área</li> <li>• Múltiples infografías</li> <li>• Tablas</li> </ul>		<ul style="list-style-type: none"> <li>• Mapa localizado de casos</li> <li>• Mapas localizadore</li> </ul>	
<b>Visualiza incertidumbre</b>		<b>Otros detalles</b>	
<b>Dirección web</b>			
<a href="https://multimedia.scmp.com/infographics/news/china/article/3047038/wuhan-virus/index.html">https://multimedia.scmp.com/infographics/news/china/article/3047038/wuhan-virus/index.html</a>			



Fuente: *South China Morning Post*.

South China Morning Post. (s.f.). Coronavirus: The disease covid-19 explained. South China Morning Post. Consultado a 8 de marzo de 2022, disponible en: <https://multimedia.scmp.com/infographics/news/china/article/3047038/wuhan-virus/index.html>

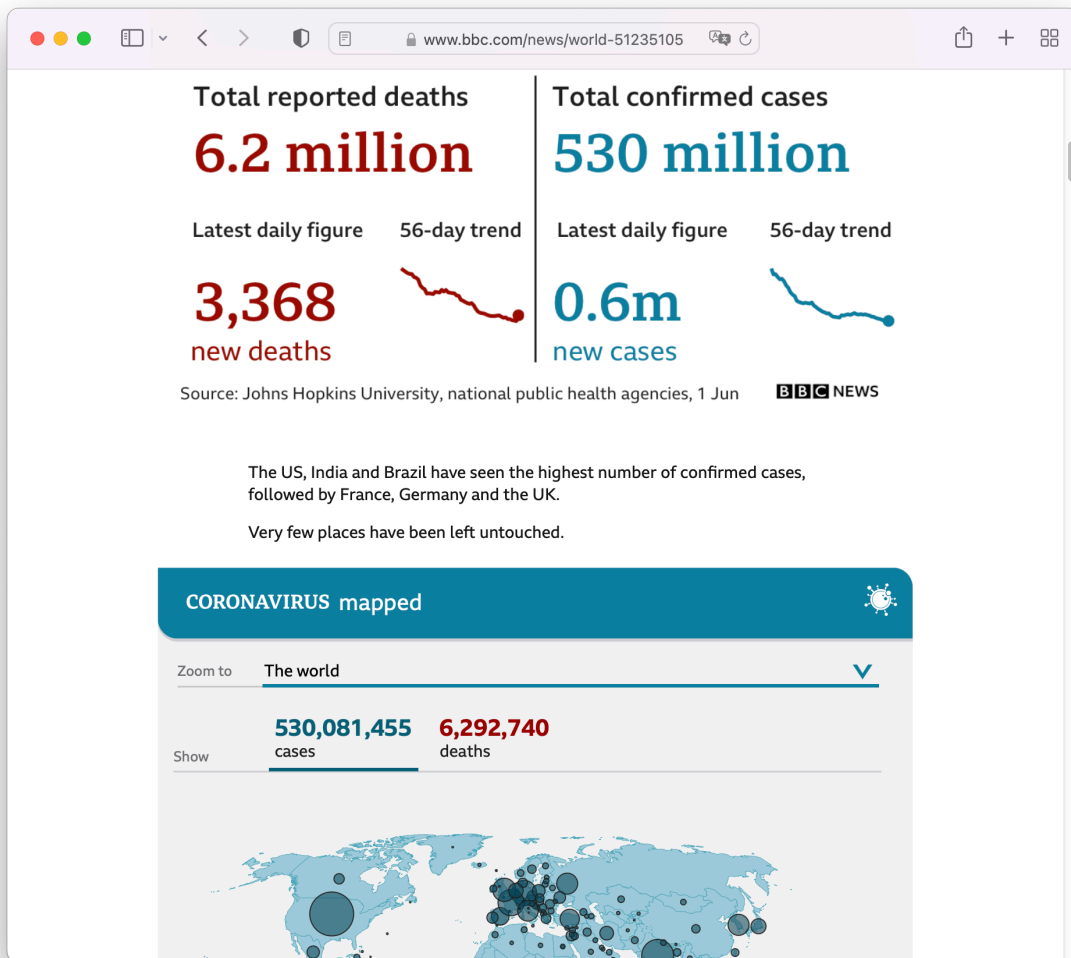
<b>Medio</b>	Straits Times	<b>País</b>	Singapur
<b>Titular</b>	How Covid-19 is spreading across the world		
<b>Descripción</b>	The coronavirus, which has infected and killed millions of people, continues to spread around the world. Here's a breakdown of the confirmed cases and deaths across Asia and the globe.		
<b>Fuentes utilizadas</b>			
JHU CSSE, MOH, Ding Xiang Yuan, Reuters		<b>Tipo de datos</b>	
		<ul style="list-style-type: none"> <li>• Casos</li> <li>• Fallecidos</li> </ul>	
<b>Tipos de gráficos</b>		<b>Tipos de mapas</b>	
<ul style="list-style-type: none"> <li>• Gráfico de barras</li> <li>• Gráfico de líneas</li> </ul>		<ul style="list-style-type: none"> <li>• Burbujas</li> </ul>	
<b>Visualiza incertidumbre</b>		<b>Otros detalles</b>	
Utiliza media semanal para ajustar los datos.			
<b>Dirección web</b>			
<a href="https://www.straitstimes.com/multimedia/graphics/2020/02/coronavirus-global-numbers/index.html">https://www.straitstimes.com/multimedia/graphics/2020/02/coronavirus-global-numbers/index.html</a>			



Fuente: *Straits Times*.

Hsien, A. L. C. (s.f.). Coronavirus: Countries and regions impacted so far. The Straits Times. Consultado a 8 de marzo de 2022, disponible en: <https://www.straitstimes.com/multimedia/graphics/2020/02/coronavirus-global-numbers/index.html>

<b>Medio</b>	BBC	<b>País</b>	Reino Unido
<b>Titular</b>	Covid map: Coronavirus cases, deaths, vaccinations by country		
<b>Descripción</b>	Covid-19 is continuing to spread around the world, with more than 530 million confirmed cases and more than six million deaths reported across almost 200 countries.		
<b>Fuentes utilizadas</b>		<b>Tipo de datos</b>	
<p>The data used on this page comes from a variety of sources. It includes figures collated by Johns Hopkins University as well as national governments and health agencies.</p> <p>When comparing figures from different countries it is important to bear in mind that not all governments are recording coronavirus cases and deaths in the same way.</p>		<ul style="list-style-type: none"> <li>• Casos</li> <li>• Fallacidos</li> <li>• Vacunas</li> </ul>	
<b>Tipos de gráficos</b>		<b>Tipos de mapas</b>	
<ul style="list-style-type: none"> <li>• Gráfico de barras</li> <li>• Gráfico de líneas</li> </ul>		<ul style="list-style-type: none"> <li>• Burbuja</li> <li>• Coropletas</li> </ul>	
<b>Visualiza incertidumbre</b>		<b>Otros detalles</b>	
Utiliza media semanal para ajustar los datos.			
<b>Dirección web</b>			
<a href="https://www.bbc.com/news/world-51235105">https://www.bbc.com/news/world-51235105</a>			



Fuente: BBC.

The Visual and Data Journalism Team. (s.f.). Covid map: Coronavirus cases, deaths, vaccinations by country. BBC News. Consultado a 8 de marzo de 2022, disponible en: <https://www.bbc.com/news/world-51235105>

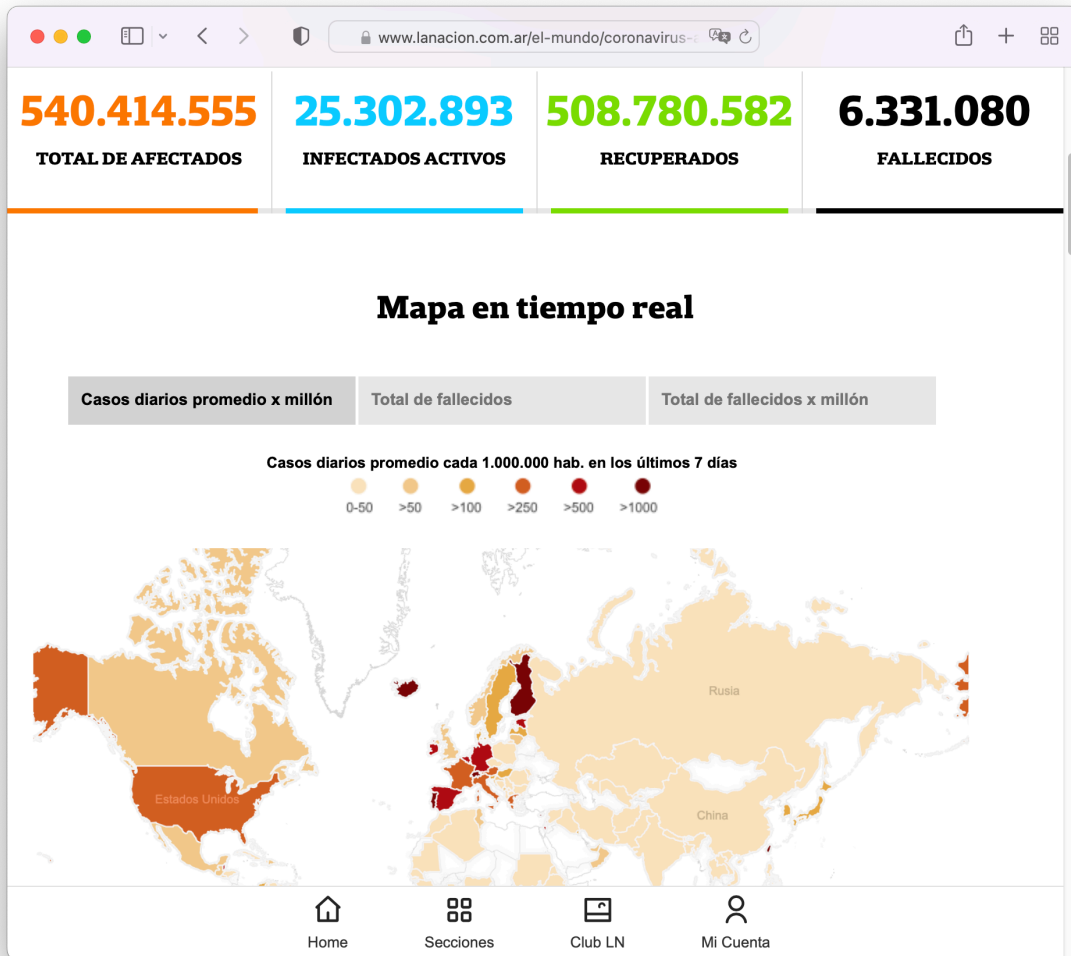
<b>Medio</b>	El País	<b>País</b>	España
<b>Titular</b>	El mapa del coronavirus en el mundo: así avanzan los contagios y las muertes día a día		
<b>Descripción</b>	Para seguir la evolución de la pandemia analizamos tres variables: la movilidad, el ritmo del crecimiento del virus (R), su incidencia –casos notificados en 14 días y por 100.000 habitantes– y el nivel de riesgo. Para definir el color de cada país usamos un semáforo: riesgo alto, cuando la incidencia es superior a 150 casos, riesgo medio, cuando supera los 50 casos y riesgo bajo cuando la incidencia es menor.		
<b>Fuentes utilizadas</b>		<b>Tipo de datos</b>	
<p>Los datos nacionales son del Center for Systems Science and Engineering (CSSE) de la Universidad Johns Hopkins. Para España usamos los datos del Ministerio de Sanidad.</p> <p>El número reproductivo R. Para calcular el número reproductivo del virus en cada momento utilizamos el paquete EpiEstim de R. Para cada país o región y día consideramos el incremento diario absoluto de casos y muertes en cada lugar de los últimos siete días. Suponemos un tiempo entre contagios medio de 4,7 días con una desviación típica de 2,9 de forma similar a la de este trabajo de la London School of Hygiene &amp; Tropical Medicine.</p> <p>Movilidad. Para medir el funcionamiento de la sociedad usamos dos fuentes de datos de movilidad agregados y anonimizados: los que ofrecen Google y Apple.</p>		<ul style="list-style-type: none"> <li>• Casos</li> <li>• Fallecidos</li> <li>• Incidencia</li> </ul>	
<b>Tipos de gráficos</b>		<b>Tipos de mapas</b>	
<ul style="list-style-type: none"> <li>• Gráfico de barras</li> <li>• Gráfico de línea</li> <li>• Gráfico de calendario</li> </ul>		<ul style="list-style-type: none"> <li>• Mapa de coropletas</li> </ul>	
<b>Visualiza incertidumbre</b>		<b>Otros detalles</b>	
Utiliza una variable como incidencia para aproximar los datos			
<b>Dirección web</b>			
<a href="https://elpais.com/especiales/coronavirus-covid-19/el-mapa-del-coronavirus-en-el-mundo/">https://elpais.com/especiales/coronavirus-covid-19/el-mapa-del-coronavirus-en-el-mundo/</a>			



Fuente: El País.

El País. (s.f.). El Mapa del coronavirus en el mundo: Así avanzan Los Contagios y las muertes día a día. EL PAÍS. Consultado a 8 de marzo de 2022, disponible en: <https://elpais.com/especiales/coronavirus-covid-19/el-mapa-del-coronavirus-en-el-mundo/>

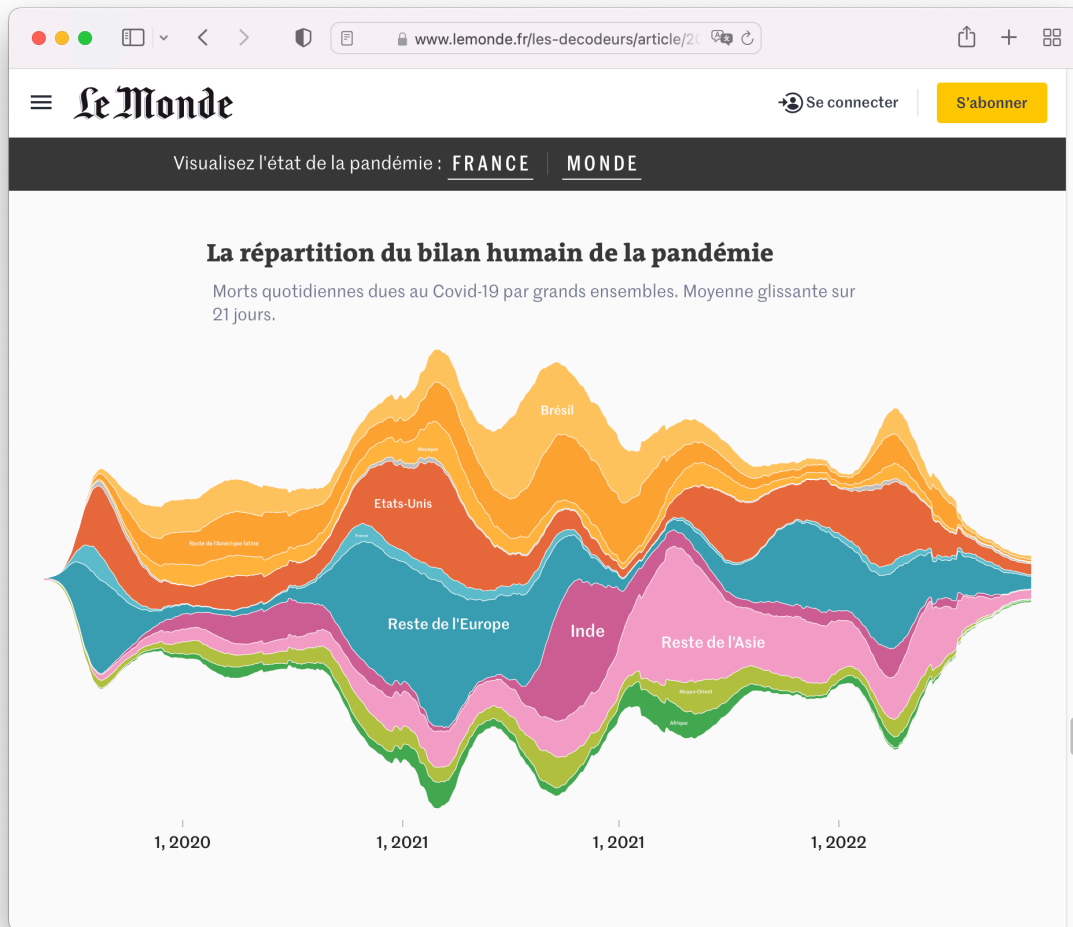
<b>Medio</b>	La Nación	<b>País</b>	Argentina
<b>Titular</b>	El avance de la pandemia en el mundo		
<b>Descripción</b>	No incluye descripción.		
<b>Fuentes utilizadas</b>			
Johns Hopkins Whiting School of Engineering en base a la OMS		<b>Tipo de datos</b>	
		<ul style="list-style-type: none"> <li>• Afectados</li> <li>• Activos</li> <li>• Recuperados</li> <li>• Fallecidos</li> </ul>	
<b>Tipos de gráficos</b>		<b>Tipos de mapas</b>	
<ul style="list-style-type: none"> <li>• Gráfico de barras</li> <li>• Gráfico de línea</li> <li>• Tablas</li> <li>• Gráfico de área</li> </ul>		<ul style="list-style-type: none"> <li>• Coropletas</li> </ul>	
<b>Visualiza incertidumbre</b>		<b>Otros detalles</b>	
Utiliza media semanal para ajustar los datos.		Ajusta por población	
<b>Dirección web</b>			
<a href="https://www.lanacion.com.ar/el-mundo/coronavirus-asi-se-propaga-virus-mundo-nid2351138/">https://www.lanacion.com.ar/el-mundo/coronavirus-asi-se-propaga-virus-mundo-nid2351138/</a>			



Fuente: La Nación.

La Nación. (s.f.). Coronavirus: Así Se propaga el virus en el mundo. LA NACION. Consultado a 8 de marzo de 2022, disponible en: <https://www.lanacion.com.ar/el-mundo/coronavirus-asi-se-propaga-virus-mundo-nid2351138/>

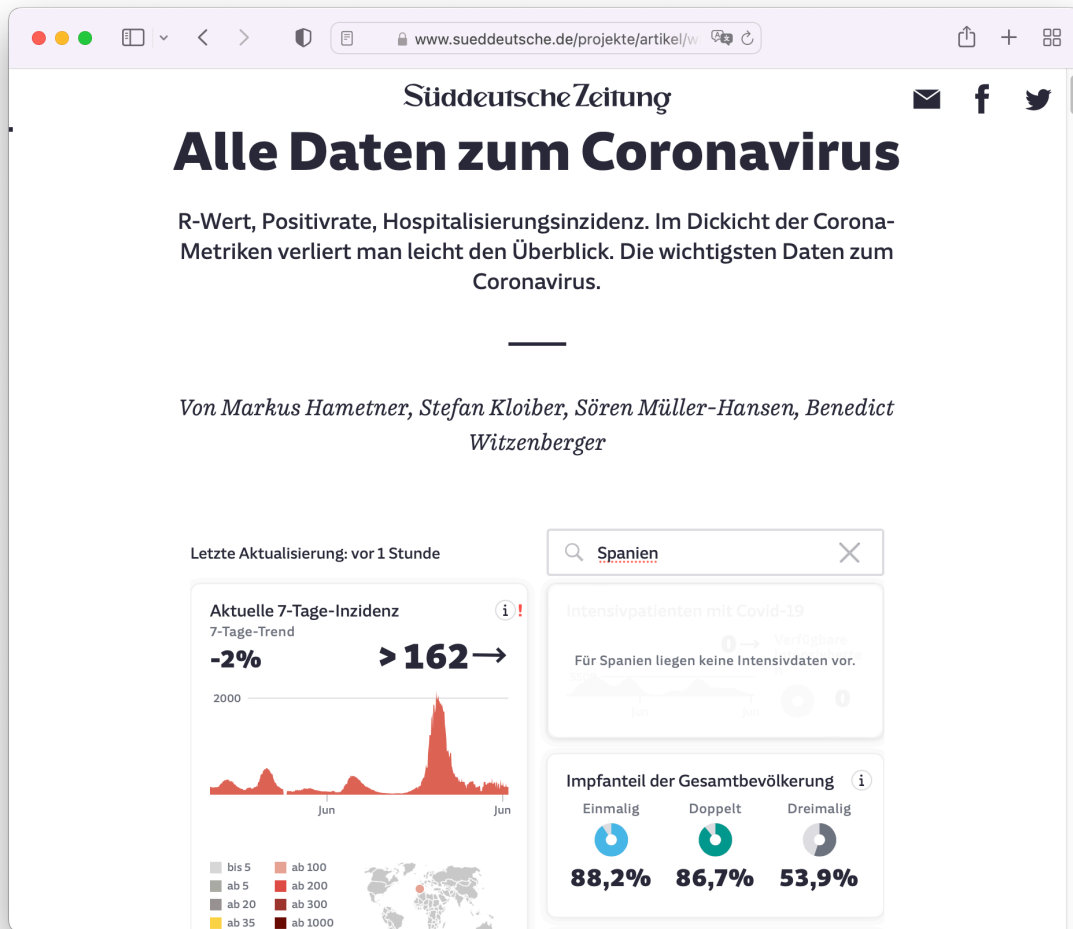
<b>Medio</b>	Le Monde	<b>País</b>	Francia
<b>Titular</b>	Covid-19 : le tableau de bord de l'épidémie		
<b>Descripción</b>	La situation sanitaire en France et à l'étranger, en carte et en graphiques, actualisée chaque jour entre 19 heures et 20 heures.		
<b>Fuentes utilizadas</b>		<b>Tipo de datos</b>	
Johns-Hopkins University		<ul style="list-style-type: none"> <li>• Casos</li> <li>• Fallecidos</li> </ul>	
<b>Tipos de gráficos</b>		<b>Tipos de mapas</b>	
<ul style="list-style-type: none"> <li>• Gráficos de barras</li> <li>• Streamgraph</li> <li>• Tablas</li> </ul>		<ul style="list-style-type: none"> <li>• Mapa de coropletas</li> <li>• Cartograma</li> </ul>	
<b>Visualiza incertidumbre</b>		<b>Otros detalles</b>	
Utiliza media semanal para ajustar los datos.			
<b>Dirección web</b>			
<a href="https://www.lemonde.fr/les-decodeurs/article/2022/06/09/covid-19-le-tableau-de-bord-de-l-epidemie_6038751_4355773.html">https://www.lemonde.fr/les-decodeurs/article/2022/06/09/covid-19-le-tableau-de-bord-de-l-epidemie_6038751_4355773.html</a>			



Fuente: Le Monde.

Baruch, J., Aubert, R., Ferrer, M., Sanchez, L., Dagorn, G. y Breteau, P. (s.f.). Covid-19 : Le Tableau De Bord de l'épidémie. Le Monde. Consultado a 8 de marzo de 2022, disponible en: [https://www.lemonde.fr/les-decodeurs/article/2022/06/09/covid-19-le-tableau-de-bord-de-l-epidemie\\_6038751\\_4355773.html](https://www.lemonde.fr/les-decodeurs/article/2022/06/09/covid-19-le-tableau-de-bord-de-l-epidemie_6038751_4355773.html)

<b>Medio</b>	Süddeutsche Zeitung	<b>País</b>	Alemania
<b>Titular</b>	Alle Daten zum Coronavirus		
<b>Descripción</b>	R-Wert, Positivrate, Hospitalisierungsinzidenz. Im Dickicht der Corona-Metriken verliert man leicht den Überblick. Die wichtigsten Daten zum Coronavirus.		
<b>Fuentes utilizadas</b>		<b>Tipo de datos</b>	
Our World in Data y Johns Hopkins University		<ul style="list-style-type: none"> <li>• Casos</li> <li>• Fallecidos</li> <li>• Hospitalizationes</li> <li>• Vacunas</li> </ul>	
<b>Tipos de gráficos</b>		<b>Tipos de mapas</b>	
<ul style="list-style-type: none"> <li>• Gráficos de barras</li> <li>• Gráfico de tartas</li> <li>• Tablas</li> <li>• Flechas</li> </ul>		<ul style="list-style-type: none"> <li>• Mapa de coropletas</li> </ul>	
<b>Visualiza incertidumbre</b>		<b>Otros detalles</b>	
Utiliza media semanal para ajustar los datos.			
<b>Dirección web</b>			
<a href="https://www.sueddeutsche.de/projekte/artikel/wissen/coronavirus-die-wichtigsten-daten-in-grafiken-e536967/">https://www.sueddeutsche.de/projekte/artikel/wissen/coronavirus-die-wichtigsten-daten-in-grafiken-e536967/</a>			



Fuente: Süddeutsche Zeitung.

Süddeutsche Zeitung. (s.f.). Alle datos zum coronavirus. Süddeutsche.de. Consultado a 8 de marzo de 2022, disponible en: <https://www.sueddeutsche.de/projekte/artikel/wissen/coronavirus-die-wichtigsten-daten-in-grafiken-e536967/>

<b>Medio</b>	The Guardian	<b>País</b>	Reino Unido
<b>Titular</b>	Covid world map: which countries have the most coronavirus vaccinations, cases and deaths?		
<b>Descripción</b>	Covid-19 has spread around the planet, sending billions into lockdown. Find out where the virus has spread, and where it is now most deadly		
<b>Fuentes utilizadas</b>		<b>Tipo de datos</b>	
Our World in Data		<ul style="list-style-type: none"> <li>• Casos</li> <li>• Fallecidos</li> <li>• Vacunas</li> <li>• Tests</li> </ul>	
<b>Tipos de gráficos</b>		<b>Tipos de mapas</b>	
<ul style="list-style-type: none"> <li>• Gráficos de barras</li> <li>• Gráfico de línea</li> <li>• Tablas</li> </ul>		<ul style="list-style-type: none"> <li>• Mapa de coropletas</li> </ul>	
<b>Visualiza incertidumbre</b>		<b>Otros detalles</b>	
Utiliza media semanal para ajustar los datos.			
<b>Dirección web</b>			
<a href="https://www.theguardian.com/world/2021/dec/01/covid-world-map-which-countries-have-the-most-coronavirus-vaccinations-cases-and-deaths">https://www.theguardian.com/world/2021/dec/01/covid-world-map-which-countries-have-the-most-coronavirus-vaccinations-cases-and-deaths</a>			

www.theguardian.com/world/2021/dec/01

**Coronavirus**

# Covid world map: which countries have the most coronavirus vaccinations, cases and deaths?

**Covid-19 has spread around the planet, sending billions into lockdown. Find out where the virus has spread, and where it is now most deadly**

- **Coronavirus map of the UK**
- **Coronavirus map of the US**
- **Coronavirus cases in Australia**

**Pablo Gutiérrez, Seán Clarke and Ashley Kirk**

Wed 1 Dec 2021 06.50 EST

f t e

## Latest global Covid-19 data

Total cases	<b>535.26m</b>	New daily cases	<b>195,695</b>
Total deaths	<b>6.31m</b>	New daily deaths	<b>455</b>
% fully vaccinated	<b>60.6%</b>	New daily vaccinations	<b>0.0m</b>

% fully vaccinated figure refers to people who have received all doses prescribed by the initial vaccination protocol, as a proportion of the total world population. Data from [JHU CSSE Covid-19 Data](#) and Our World in Data at 03.02 on 13 June 2022

Fuente: The Guardian.

The Guardian. (s.f.). Covid World Map: Which countries have the most coronavirus vaccinations, cases and deaths? The Guardian. Consultado a 8 de marzo de 2022, disponible en: <https://www.theguardian.com/world/2021/dec/01/covid-world-map-which-countries-have-the-most-coronavirus-vaccinations-cases-and-deaths>



## **Anexo IV. Selección de bots con fines periodísticos analizados en esta tesis**

A continuación se reproducen los bots seleccionados en esta tesis para estudiar su uso en el sector del periodismo. La clasificación de bots con diferentes aplicaciones en el sector del periodismo propuesta en esta tesis se ha elaborado tras analizar de forma minuciosa cada una de las características de los bots enumerados en este anexo. La base de datos completa, con todos los detalles y características de los bots analizados, puede consultarse en el siguiente enlace:

[https://docs.google.com/spreadsheets/d/1\\_CzxZ1kGOw\\_T1whJ632N4v8DiduJPHHvqKx6IybCzX0/edit?usp=sharing](https://docs.google.com/spreadsheets/d/1_CzxZ1kGOw_T1whJ632N4v8DiduJPHHvqKx6IybCzX0/edit?usp=sharing)

La lista de bots analizados sigue a continuación:

- BBG Cherry Blossoms, <https://twitter.com/bbgblossoms>
- censusAmericans, <https://twitter.com/censusAmericans>
- colorschemez, <https://twitter.com/colorschemez>
- Every tract, <https://twitter.com/everytract>
- California Wildfires map, <https://www.latimes.com/wildfires-map/>
- FiveyFox, <https://fivethirtyeight.com/contributors/fivey-fox/>
- FOIAMachine, <https://www.foiamachine.org/>
- ImmigrantsAreUS, <https://twitter.com/ImmigrantsAreUS>
- Klaxon, <https://newsklaxon.org/>

- Michelle Singletary's digital equivalent, <https://www.washingtonpost.com/graphics/2020/business/retirement-planning-bot/>
- newgovrepos, <https://twitter.com/newgovrepos>
- News Homepages, <https://twitter.com/newshomepages>
- newsdiffs, <https://twitter.com/newsdiffs>
- newsnerdrepos, <https://twitter.com/newsnerdrepos>
- NYT\_first\_said, [https://twitter.com/NYT\\_first\\_said](https://twitter.com/NYT_first_said)
- NYTHaikkus, <https://twitter.com/nythaikus>
- Politibot, <https://twitter.com/politibot>
- Quakebot, <https://www.latimes.com/people/quakebot>
- reddit place bot, <https://github.com/martgnz/reddit-place-bot>
- Trending Notebook, <https://twitter.com/trendingnotebo2>
- TweteoSwiss, <https://twitter.com/TweteoSwiss>
- US Air Quality Bot, [https://twitter.com/air\\_qual\\_bot\\_us](https://twitter.com/air_qual_bot_us)
- Weather Warning Maps, [https://twitter.com/w\\_w\\_maps](https://twitter.com/w_w_maps)
- WordleBot, <https://www.nytimes.com/interactive/2022/upshot/wordle-bot.html>

- World Development Indicators (bot), [https://twitter.com/air\\_qual\\_bot\\_us](https://twitter.com/air_qual_bot_us)
- A.I. drawing bot, <https://pudding.cool/projects/caption-contest/>
- Heat records, <https://pudding.cool/projects/heat-records/>

