

Transparencia, conservación y paridad en los grandes proyectos de datos: el caso de los Data Journalism Awards (DJA) y Sigma Awards (2013-2020)

(Transparency, conservation and parity in big data projects: the case of the Data Journalism Awards (DJA) and Sigma Awards (2013-2020))

Abellán Mancheño, Andrea¹; Arias Robles, Félix²; García-Avilés, José Alberto³

Universidad Miguel Hernández de Elche (España)
abellanandrea8@gmail.com¹; farias@umh.es²;
jose.garciaa@umh.es³

Recep.:18.02.2022

BIBLID [eISSN 1988-3935 (2022), 19; 107-124]

Acep.: 14.04.2022

El periodismo de datos ha crecido exponencialmente, convirtiéndose en referencia informativa. Esta investigación explora tres características fundamentales de esta disciplina a partir del análisis de proyectos de primer nivel: 1) la transparencia de sus proyectos a través del análisis de su metodología; 2) la paridad en los equipos; y 3) su conservación, pues muchas de las piezas se publican solo en formato digital, lo que dificulta su pervivencia. La investigación se centra en el análisis de contenido de los proyectos galardonados por los Data Journalism Awards (DJA) desde 2013 hasta 2019 y por los Sigma Awards en 2020. El estudio aborda las mejores visualizaciones, mejores apps e investigaciones de datos. Los resultados reflejan que las visualizaciones son los proyectos que gozan de una mejor pervivencia, que las metodologías se comparten más en las investigaciones con datos y que la presencia de mujeres no supera el 50% en ninguna de las categorías galardonadas.

Palabras clave: Periodismo de datos. Transparencia. Conservación. Paridad. Data Journalism Awards. Sigma Awards.

Datuen kazetaritza esponentzialki hazi da, informazio-erreferentzia bihurtuz. Ikerketa honek diziplina honen funtsezko hiru ezaugarri aztertzen ditu, lehen mailako proiektuen azterketatik abiatuta: 1) bere proiektuen gardentasuna bere metodologiaren analisiaren bidez; 2) ekipoen parekotasuna; eta 3) bere kontserbazioa, pieza asko formatu digitalean bakarrik argitaratzen baitira, eta horrek zaildu egiten du bere iraupena. Ikerketaren ardatza Data Journalism Awards (DJA) sariak 2013tik 2019ra eta Sigma Awards sariak 2020an saritutako proiektuen edukiaren analisia da. Azterlanak bistaratze onenak, app onenak eta datuen ikerketak jorratzen ditu. Emaitzek erakusten dute bistaratzeak direla biziraupen onena duten proiektuak, metodologiak gehiago partekatzen direla ikerketetan datuekin, eta emakumeen presentzia ez dela % 50etik gorakoa saritutako kategoria bakar batean ere.

Gako hitzak: Datuen kazetaritza. Gardentasuna. Kontserbazioa. Parekotasuna, Data Journalism Awards. Sigma Awards.

Data journalism has grown exponentially to become an informative reference. This study examines three of the most important features of this discipline through the analysis of high-level projects: 1) the transparency of the projects through the analysis of their methodology; 2) the parity in the teams; and 3) their preservation, as many of the works are published only in digital form, making their survival difficult. The study focuses on the content analysis of the projects awarded by the Data Journalism Awards (DJA) from 2013 to 2019 and by the Sigma Awards in 2020. The study focuses on the best visualisations, best apps, and data investigations. The results show that visualisations are the projects that survive best, that methodologies are more shared in data investigations, and that the presence of women does not exceed 50% in any of the award-winning categories.

Keywords: Data journalism. Transparency. Curation. Parity. Data Journalism Awards. Data Journalism Awards. Sigma Awards.

Le journalisme de données (ou data journalisme) s'est développé de manière exponentielle pour devenir une référence de l'information. Cette étude explore trois caractéristiques fondamentales de cette discipline en partant de l'analyse de projets de premier ordre : 1) la transparence de ces projets à travers l'analyse de leur méthodologie ; 2) la parité des équipes ; et 3) leur conservation, puisque, la majorité des pièces étant publiées uniquement en format numérique, il est difficile d'assurer leur pérennité. L'étude est centrée sur l'analyse de contenu des projets récompensés par les Data Journalism Awards (DJA) de 2013 à 2019 et par les Sigma Awards en 2020, en abordant les meilleures visualisations, les meilleures applications et recherches de données. Les résultats montrent que les visualisations sont les projets qui perdurent le plus, que les méthodologies se partagent davantage dans les recherches avec des données et que la présence des femmes est inférieure à 50 % dans toutes les catégories primées.

Mots clés: Journalism de données. Data journalisme. Transparence. Conservation. Parité. Data Journalism Awards. Sigma Awards.

1. INTRODUCCIÓN

El periodismo de datos es una de las disciplinas periodísticas sobre las que más se ha discutido y escrito a principios del siglo XXI, tanto en el discurso profesional como en el académico (Mair, Radcliffe y Felle, 2015). La disciplina ha crecido hasta abarcar muchas técnicas y métodos distintos tales como la visualización de datos, el uso de bases de datos o las noticias interactivas (Anderson, 2018). Estos formatos suelen publicarse en línea, relegando el contenido en papel a un segundo plano. Pero la naturaleza dinámica de la web y de las plataformas de distribución digitales, que están en constante cambio, hace que la pervivencia de estos proyectos sea complicada (Broussard y Boss, 2018). De hecho, noticias que han cubierto sucesos históricos como los atentados del 11 de septiembre ya no están disponibles. Y es que “el periodismo se considera a menudo el primer borrador de la historia, pero ¿qué ocurre cuando ese borrador se escribe en un programa informático que se queda obsoleto?” (Duffy y Flynn, 2021).

El periodismo de datos ha ganado notable importancia en los últimos años. Sucesos globales como la pandemia provocada por el virus de la COVID-19 han demostrado la importancia en el uso de los datos en los trabajos periodísticos y la necesidad fundamental de transmitirlos de forma clara y efectiva a una audiencia ávida de información (García Avilés, 2021).

Esta investigación explora las características del periodismo de datos de alto nivel. Para ello, toma como muestra los proyectos galardonados por los DJA (*Data Journalism Awards*) y *Sigma Awards* desde 2013 hasta 2020. Se centra en tres categorías: mejor visualización, mejor app y mejor investigación de datos. En la muestra se exploran tres aspectos concretos: conservación, transparencia y paridad en los equipos de trabajo.

Teniendo en cuenta este contexto, se plantean las siguientes preguntas de investigación:

PI1: ¿Cuántos proyectos de periodismo de datos están en funcionamiento? ¿Qué elementos clave se han perdido?

PI2: ¿Comparten los proyectos galardonados su metodología y las bases de datos utilizadas?

PI3: ¿Cómo ha evolucionado la composición de los equipos de datos dentro de las redacciones? ¿Ha aumentado o disminuido el número de mujeres que los conforman?

2. MARCO TEÓRICO

El periodismo de datos puede entenderse como el intento de los medios de comunicación de adaptarse y responder a los cambios en nuestro entorno informativo, incluyendo una narración multidimensional más interactiva que permita a los lectores explorar las fuentes subyacentes a las noticias y los anime a participar en el proceso de creación y evaluación de las historias (Viana, 2012). Aaron Pilhofer (2012) hace hincapié en las múltiples opciones que aglutina esta disciplina que van desde la información asistida por ordenador (que utiliza los datos como "fuente") hasta las visualizaciones de datos y las aplicaciones de noticias más vanguardistas. El objetivo común es el periodístico: proporcionar información y análisis que permitan recabar información fiable sobre los temas importantes de la actualidad.

Los comienzos del periodismo de datos van acompañados con la evolución de conceptos y términos básicos forjados a principios de los años setenta. Se trata del CAR, (*Computer-Assisted Reporting*) o periodismo asistido por ordenador, al que se fueron sumando una amplia variedad de tecnologías y herramientas propias de la era digital (Mair; Radcliffe, y Felle, 2015). En el contexto actual, en el que cada vez se recopilan más datos y hay más herramientas disponibles, el vínculo entre los datos y el periodismo es cada vez más fuerte (Gray, Bounegru y Chambers, 2012). La creciente importancia del periodismo de datos radica en la capacidad de los profesionales para proporcionar un mayor contexto, claridad y credibilidad al sobrecargado panorama digital y en el desarrollo de procesos innovadores y nuevas narrativas que permitan reconectar con las audiencias (García-Avilés, 2020).

Cada vez son más los proyectos de datos que se publican únicamente en el entorno digital, cuya pervivencia no está garantizada. Rothenberg (2015) identifica cuatro retos que dificultan la conservación de los archivos digitales: el deterioro físico de los soportes, la pérdida de información sobre el formato, la codificación o la compresión de los archivos, la obsolescencia del hardware y la falta de disponibilidad de este. En los últimos años, numerosos autores han estudiado las opciones que garantizan la conservación de los proyectos periodísticos. Entre ellas, el uso de herramientas tales como el rastreador web *Heritrix* de Internet Archive (Broussard, 2015a).

Sin embargo, estas soluciones no son perfectas, y, por ejemplo, el contenido publicado a través de muros de suscripción o que requiere un registro por parte del usuario, no se almacena de forma automática (Broussard, 2021). Además, no sirven para capturar contenido dinámico, como los *feeds* de las redes sociales o las visualizaciones interactivas (Hansen y Paul 2015; Hansen y Paul 2017). Aunque sí existen herramientas enfocadas a preservar el contenido dinámico, como *Webrecorder* desarrollado en JavaScript, no funcionan de manera automática, y necesitan de la intervención del usuario para ser ejecutadas (Broussard, 2018). Además, la utilización de estos recursos todavía no es habitual, Meredith Broussard y

Katherine Boss entrevistaron a 76 periodistas de 41 medios de comunicación sobre cuántos de estos medios tenían un sistema de conservación para las aplicaciones de noticias, y, solo el 5% afirmaron utilizarlos. Este estudio averigua si esa cifra se mantiene o no en la muestra seleccionada. Con este fin, se investiga si es posible todavía acceder a los proyectos seleccionados y cuáles son las condiciones que interfieren en su mantenimiento o desaparición (Boss y Broussard, 2018).

El periodismo de datos ofrece otras oportunidades, tales como aumentar la transparencia y la reproducibilidad de los proyectos. A causa de la democratización de los datos (Mair; Radcliffe, y Felle, 2015), la información ha dejado de pertenecer solo a los expertos, lo que ha hecho del periodismo de datos una pieza clave para abanderar la reivindicación de la transparencia informativa. Del mismo modo, el número de redacciones que ponen sus bases de datos y el código de sus proyectos a disposición del público también va en aumento y la atención se ha desplazado de los datos al proceso de gestión y presentación de estos (Mair; Radcliffe, y Felle, 2015). Para Chery Phillips, las razones de compartir las bases de datos implicadas en un proyecto con la audiencia son las siguientes: aumentar la transparencia y permitir a los lectores los utilicen, así como facilitar feedback sobre el uso que se ha hecho de los mismos (Phillips, 2012). Dar acceso a los apéndices metodológicos ayuda a que el código y los datos utilizados en una pieza puedan ser reutilizados y que, además, la audiencia pueda identificar posibles errores y sesgos en su uso (Broussard, 2015b).

La desigualdad de género no es exclusiva del periodismo de datos ni del periodismo en general. Pero, a medida que la comunidad que gira en torno al periodismo de datos crece, también debería hacerlo el interés por conseguir que los equipos que lo abanderan sean igualitarios (Alexander, 2016). El estudio de la paridad dentro del periodismo de datos resulta de especial interés, pues es una disciplina que se nutre de la intersección de la tecnología y el periodismo, exigiendo a los periodistas conocimientos de programación muy especializados en comparación con otros cometidos (Alexander, 2016). Según algunos expertos (Shoemaker y Reese, 1996), existe una conexión entre las características individuales, como el género de los periodistas, y el contenido que producen. Dichos autores argumentan que las reporteras mujeres son más propensas que los hombres a utilizar fuentes femeninas en sus historias y a dar más importancia a esas fuentes, lo que tiene un impacto en la inclusividad de la información que transmiten (Cassidy, 2008).

3. METODOLOGÍA

Esta investigación explora la evolución del periodismo de datos desde 2013. Para ello, se ha llevado a cabo un estudio cualitativo de tres categorías de los proyectos galardonados por los *Data Journalism Awards* (DJA),

gestionados por la *Global Editors' Network* (GEN) desde 2013 hasta 2019 y por los *Sigma Awards* en su primera edición en 2020.

Los *Data Journalism Awards* (DJA), organizados por la *Global Editors Network* desde 2012, supusieron “la creación de los primeros galardones internacionales que reconocen los trabajos sobresalientes de periodismo de datos a nivel mundial” (Córdoba-Cabús, 2020). Creados por Simon Roger, el jurado siempre ha estado compuesto por periodistas de datos de alto nivel. Paul Steiger, Paul Radu, Peter Barron, Reginald Chua y Wolfgang Blau son algunos de los expertos que han evaluado los proyectos a lo largo de estos años. A partir de 2020, y a consecuencia de la desaparición de la organización que los fundó, pasaron a ser conocidos como *Sigma Awards*, lo que supuso algo más que un cambio de nombre según Reginald Chua, copresidente del proyecto: “No queríamos que los galardones desaparecieran, sino que se convirtieran en algo completamente nuevo” (Chua, 2020). Los premios tienen como objetivo “destacar el mejor periodismo de datos que se hace en todo el mundo; crear programas y recursos en torno a los premios que permitan a personas de dentro y fuera de la comunidad del periodismo de datos aprender de este trabajo; utilizar los premios como una forma de unir, galvanizar y ampliar las comunidades de periodismo de datos en todo el mundo” (Chua, 2020). Desde 2020, Simon Rogers es su director, Paul Steiger se incorporó como asesor y Marianne Bouchart gestiona los premios con la ayuda de Kuek Ser Kuang Keng.

Estos galardones reconocen cada año los mejores proyectos centrados en datos, “valorando la calidad e innovación en el uso y el análisis de datos, el contenido y la presentación de la información” (Córdoba-Cabús, 2020). Diversos investigadores han estudiado los proyectos galardonados en este certamen desde varias perspectivas. Córdoba-Cabús (2020) exploró las características de los proyectos premiados en 2019 a partir de sus temáticas, las fuentes de datos utilizadas y el uso de visualizaciones. Por su parte, Chaparro-Domínguez y Díaz Campo (2021) analizaron los proyectos premiados por los *Data Journalism Awards*, *Sigma Awards* y *Online Journalism Awards* (OJA) desde 2012 hasta 2020 para discernir los estándares éticos de los proyectos reconocidos, teniendo en cuenta la verificación y análisis de los datos, la transparencia y la privacidad. Por otro lado, Loosen, Reimer y De Silva-Schmidt (2017) estudiaron las características de los proyectos nominados a los *Data Journalism Awards* entre 2013 y 2016 con objeto de comprobar si sus características han cambiado entre esos años, analizando los equipos que los elaboraron, los temas que cubrían, qué tipo de fuentes de datos utilizaban y qué visualizaciones o elementos interactivos incluían.

Teniendo en cuenta los estudios previos, este trabajo da respuesta a los vacíos de investigación existentes sobre el periodismo de datos de primer nivel. Se centra en investigar sobre la transparencia, la paridad y la pervivencia de los proyectos. Aunque la estructura y las secciones de los premios han cambiado cada año, hay cinco categorías que se han mantenido estables: mejor visualización, mejor investigación con datos, mejor app, mejor

página web sobre periodismo de datos y mejor portfolio. Por su relevancia dentro del periodismo de datos y con el fin de acotar la muestra, en este estudio se tienen en cuenta los proyectos reconocidos en tres categorías: mejor visualización de datos, mejor investigación basada en datos y mejor aplicación. No se ha tenido en cuenta la última edición de los *Sigma Awards* en 2021 puesto que los premios no se diferenciaron por categorías, y la ausencia de este parámetro hubiera alterado el corpus del estudio.

En esta investigación se ha realizado un análisis de contenido de las páginas web de los proyectos galardonados. Este método es destacado por Krippendorff (2013) por su capacidad para hacer inferencias replicables y válidas a partir de los textos (u otras materias significativas, en este caso páginas webs) al contexto de su uso. En este caso, primero hemos analizado las páginas web en las que los proyectos ganadores se anunciaron, y luego, cada una de las páginas dónde los proyectos se alojaron.

Las visualizaciones son uno de los elementos diferenciadores del periodismo de datos que permite que conceptos abstractos e invisibles se conviertan en una "imagen" visible en nuestra pantalla (Gray; Bounegru y Chambers, 2012). Numerosos autores reconocen el papel de las visualizaciones, ya sean estáticas o interactivas, en los procesos previos y posteriores a la publicación de un proyecto. Para ellos, durante la etapa de desarrollo pueden servir para diferenciar parámetros, temas e incluso errores en una base de datos, mientras que en el momento de la publicación contribuyen a ilustrar los aspectos más importantes de la noticia de una forma más atractiva, eliminando la información demasiado técnica (Gray; Bounegru y Chambers, 2012).

La segunda categoría analizada tiene en cuenta las investigaciones basadas en datos. Éstas son fundamentales, ya que, cada vez hay más bases de datos abiertos disponibles. El periodismo de datos y como consecuencia, las investigaciones que surgen de él, "aprovecha de un volumen de datos inmenso y que aumenta continuamente, lo que se conoce como *Big Data* o *grandes datos*, en referencia a los sistemas de información y comunicación que gestionan grandes conjuntos de datos o *data sets*. Estos datos son más o menos públicos y más o menos accesibles a través de la web -bien sea en el propio contenido, ayudado por el *HTML* o en los documentos enlazados-, en bases de datos o a través de otras herramientas informáticas de búsqueda. Luego hay que filtrar, seleccionar y visualizar de diversas formas posibles, en formato papel o audiovisual, vía web o TV" (Bravo, 2013: 102).

Por último, se describen las aplicaciones periodísticas premiadas. Cuando se orientan a fines periodísticos, las aplicaciones permiten simplificar información compleja para atraer a aquellas audiencias con menos tiempo o que simplemente no eligen informarse a través de piezas en profundidad (Gray; Bounegru y Chambers, 2012). Estos formatos interactivos o "piezas de software diseñadas para contar una historia" (Broussard, 2015a) ayudan a los lectores a encontrar su conexión personal con un fenómeno amplio,

permitiéndole relacionarlo con su conocimiento previo, lo que favorece una comprensión más profunda. Las aplicaciones periodísticas se conforman de los siguientes componentes: una base de datos, los datos de la base de datos, la interfaz gráfica única que aparece en el navegador, a través de la cual el usuario interactúa con la base de datos, una o varias historias basadas en texto, fotos o ilustraciones (Broussard, 2015a).

Bajo estas premisas, se han considerado las siguientes variables:

1. Nombre del proyecto
2. Categoría en la que ha sido reconocido
3. Año en el que ha sido premiado
4. Estado de conservación: es accesible online o no
5. Comparte su metodología
6. Medio en el que ha sido publicado
7. Equipo premiado: número total de integrantes
8. Porcentaje de mujeres en el equipo
9. Enlace al proyecto

En total, se han analizado 35 proyectos que han sido premiados desde 2013 en las categorías mencionadas. En 2013 no se otorgó ningún premio a la mejor visualización, así que esta categoría no se ha tenido en cuenta en ese año. Tres proyectos ya no están disponibles en línea, por lo que con respecto a ellos no hemos podido obtener información acerca de los equipos que los diseñaron y si la metodología estaba o no disponible.

En el análisis de la paridad no se han tenido en cuenta los datos de equipos de tres proyectos premiados como mejor investigación de datos: *Swiss Leaks: Murky Cash*; *Luxembourg Leaks: Global Companies' Secrets Exposed* y *The Panama Papers*, que fueron galardonados en 2015 y 2016. Estas investigaciones fueron impulsadas por el ICIJ (*International Consortium of Investigative Journalism*) y son fruto de la colaboración entre redacciones de todo el mundo, por lo que no es posible analizar su composición de hombres y mujeres. En el estudio de paridad tampoco se ha incluido *Electionland*, premiada como mejor app en 2017, que fue resultado de la colaboración entre varias organizaciones: *ProPublica*, *Google News Lab*, *First Draft*, 14 universidades norteamericanas y múltiples redacciones. Además, más de 650 estudiantes contribuyeron con el que se ha definido como “la mayor iniciativa de colaboración informativa en un solo día de la historia” (Lord, 2016). Por el gran número de personas involucradas, y la dificultad para acceder a sus identidades, decidimos dejarlos fuera para no introducir sesgos en el estudio. También es importante aclarar que hay un mayor número de

proyectos premiados en la categoría de mejor investigación porque los premios se otorgaron a dos subgrupos, *large newsroom* y *small newsroom*, según el tamaño de las redacciones en las que se desarrollaron. Ocurre lo mismo con la categoría de mejor visualización, en 2015 y 2020 cuando se otorgaron dos premios acordes a la dimensión de la redacción: *large newsroom* y *small newsroom*. También en la categoría de mejor aplicación en 2013, 2015, y 2016 se galardonaron dos propuestas distintas para *small* y *large newsroom*.

4. RESULTADOS

En el análisis se han tenido en cuenta 35 proyectos: 9 galardonados como mejor visualización, 11 en la categoría de mejor aplicación y 15 premiados como mejor investigación con datos. En la Tabla 1 se observan los resultados divididos por categorías y años. En cada uno de ellos se han tenido en cuenta las cifras por número y porcentaje. Se han marcado con asterisco (en este orden: *Electionland*, *Swiss/Luxembourg Leaks* y *Panama Papers*) los proyectos que no se han incluido en el análisis de paridad.

Tabla 1. Análisis de la conservación, transparencia y paridad de los proyectos por categorías

Categoría	Año	Conservación		Metodología		Paridad		
		Accesibles (N)	%	Disponible (N)	%	Autores	Mujeres (N)	%
Mejor visualización	2013	-	-	-	-	-	-	-
	2014	1	100%	0	0%	8	1	12,6%
	2015	2	100%	0	0%	10	4	40%
	2016	1	100%	0	0%	2	0	0%
	2017	1	100%	1	100%	2	0	0%
	2018	1	100%	0	0%	2	1	33,30%
	2019	1	100%	0	0%	4	1	25%
	2020	2	100%	1	50%	13	6	46,15%
	Total	9	100%	2	22,22%	41	13	31,07%
Mejor aplicación	2013	2	50%	0	0%	3	1	33,33%
	2014	1	100%	0	0%	8	5	63%
	2015	2	50%	0	0%	8	2	25%
	2016	2	100%	0	0%	19	9	47,36%
	2017*	1	100%	1	100%	-	-	-
	2018	1	100%	1	100%	4	2	50%
	2019	1	100%	1	100%	1	1	100%
	2020	1	100%	1	100%	4	3	75%
	Total	11	81,80%	4	44,44%	47	23	48,90%

Mejor investigación	2013	2	100%	1	50%	6	4	66,66%
	2014	2	100%	1	50%	7	1	14,28%
	2015*	3	100%	2	66,66%	5	2	28,57%
	2016*	2	100%	2	100%	6	1	16,6%
	2017	1	100%	0	0%	5	3	60%
	2018	1	100%	1	100%	2	0	0%
	2019	1	100%	1	100%	51	26	37,50%
	2020	3	66,66%	1	50%	30	7	23,33%
	Total	15	93,30%	9	64,28%	112	44	39,28%

Fuente: Autores.

4.1. Conservación

En la categoría de mejor visualización, todos los proyectos son visibles actualmente en la red. En cuanto a las mejores apps, el porcentaje disminuye hasta el 81,80%, pues dos de ellas ya no están disponibles. En la categoría de investigación con datos, el porcentaje es del 93,30%, solo un proyecto ha dejado ya de funcionar.

Imagen 1



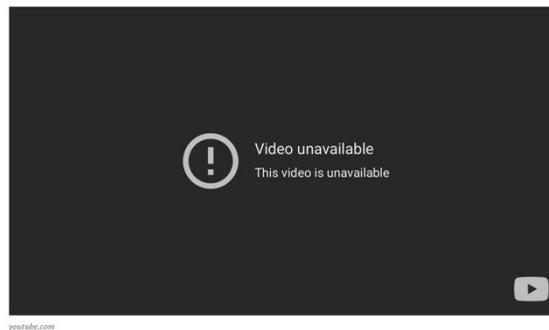
Captura del proyecto *THE ROHINGYA CRISIS- Life in the camps* de Reuters Graphics galardonado como mejor visualización en 2014. Fuente: <https://www.buzzfeednews.com/article/peteraldous/spies-in-the-skies>

Las visualizaciones son las que gozan de una mejor pervivencia, todas las consideradas en nuestra muestra siguen estando disponibles en las webs originales en la que se publicaron. Por detrás de ellas, están las investigaciones de datos. Sorprende que el único proyecto que no está disponible en esta categoría sea *Copy, Paste, Legislate* de USA Today, que fue condecorado en 2020. Así que, en este caso, no hay correlación entre el paso del tiempo y su mejor o peor conservación. Por otro lado, de un total de 11

aplicaciones investigadas hay dos que no se han podido encontrar en línea: *Great British Class Calculator*, un proyecto de la BBC premiado en 2013, y *Does school pay off? How much?* del diario mexicano El Financiero que fue reconocido en 2015.

En el estudio de esta variable, hay un factor importante a tener en cuenta, y es que, aunque los proyectos estén disponibles, en ocasiones los elementos multimedia que los complementan no lo están. Ocurre así en una de las visualizaciones premiadas, *Spies in the Skies* de BuzzFeed, galardonada en 2016. En este caso, aunque el enlace principal sí funciona, hay un vídeo alojado en YouTube que está inaccesible. Esto prueba las teorías investigadas de Meredith Broussard, que argumentan que las noticias digitales no tienen los mismos límites claramente delineados como sus contrapartes impresas “las noticias digitales están entrelazadas con el resto de Internet de maneras complejas, como cuando contienen imágenes y videos que se sirven de otros sitios web o que se extraen de una API externa como Google Maps” (Broussard, 2015a).

Imagen 2



Captura del proyecto *Spies in the Skies* de BuzzFeed publicado en <https://www.buzzfeednews.com/article/peteraldhous/spies-in-the-skies>

En el análisis de la pervivencia destaca el caso de proyectos como *Electionland*, que a día de hoy se sigue actualizando. En este caso, la versión original que fue premiada se ha perdido, pero el proyecto sí se mantiene en línea con datos nuevos. Ocurre lo mismo con la plataforma Declaraciones Juradas de La Nación que está al día desde 2015.

4.2. Transparencia en los contenidos / metodología

Para discernir sobre la transparencia en el periodismo de datos de alto nivel (PI2), se investiga cuántos proyectos del corpus comparten la metodología. Publicar un proyecto junto a su código y las fuentes de datos

utilizadas permitirá su replicación contribuyendo a mejorar el proceso científico que los rodea (Stoden y Miguez, 2014).

En la categoría de mejor visualización, solo dos proyectos la incorporaron una descripción de la metodología, el 22,22%. En el caso de las aplicaciones, son cuatro las que iban acompañadas de una metodología, el 44,44%. Por otro lado, 9 de las investigaciones premiadas si la incluían, lo que es igual al 64,28%.

Son las investigaciones de datos en las que esta práctica es más común. El 64,28% sí la incluyen, un porcentaje mayor que en el resto de las categorías. Tiene sentido pues tradicionalmente son los proyectos en los que se ha hecho un uso intensivo de datos en los que la metodología puede tomar mayor importancia (Maier, 2000) para favorecer la búsqueda de errores y garantizar su reproducibilidad. Las visualizaciones son el grupo en el que esta práctica es menos habitual, solo dos proyectos la comparten. En la categoría de mejor aplicación, se observa una tendencia al alza con el paso del tiempo, todas a partir de 2017 la incorporan.

En cuanto al modo de incluir el apéndice metodológico, muchos de los proyectos cumplen con el formato tradicional y añaden un recuadro con la información al final de la pieza (conocido como “nerd box”) que explica la metodología del proyecto (Maier, 2000). El caso de *Electionland* es una excepción también en este sentido. Dada su envergadura, se adjuntó un informe especial, en que se detalla los nombres de cada uno de los participantes y el proceso de confección del proyecto (Lord, 2016). Ocurre igual con los proyectos transnacionales mencionados por el ICIJ, en los que se ha dedicado un apartado específico en sus webs explicando cómo se llevaron a cabo y quiénes participaron en ellos.

Imagen 3

How WSJ Used an Algorithm to Analyze ‘Hamilton’ the Musical

A behind-the-scenes look at how The Wall Street Journal's Graphics team created an algorithm to help analyze the blockbuster hip-hop musical

By [Joel Eastwood](#) and [Erik Hinton](#)

Last updated June 7, 2016 at 11:2 p.m. ET | Published June 6, 2016 at 7:30 a.m. ET

“Hamilton” is the massively successful Broadway musical about America’s founding father Alexander Hamilton. By combining hip-hop and musical theatre, its writer, creator and star Lin-Manuel Miranda devised a juggernaut that found wide appeal among audiences.

But what exactly makes his lyrical construction special, and how did Mr. Miranda weave four decades of hip-hop history with three centuries of the American story? To answer part of this question, we looked to the rhyme patterns in “Hamilton,” trying to classify the specific ways Mr. Miranda references foundational rappers such as Big Pun, Rakim, and Lauryn Hill into his verses about the Revolutionary War and constitutional congress.

See the full project: [We Wrote an Algorithm to Unravel the Rhymes of Hit Musical ‘Hamilton’](#)

Ejemplo de la metodología incluida en el proyecto *The Rhymes Behind Hamilton* de *The Wall Street Journal*, galardonado en 2017 como mejor visualización:

<http://graphics.wsj.com/hamilton-methodology>

4.3. Paridad / presencia femenina

Son las aplicaciones de noticias en las que hay mayor número de mujeres, con un 48,9% del total. En la categoría de mejor visualización es donde las mujeres están menos representadas, sumando el 31,07% del total. En ninguna de las disciplinas las mujeres superan el 50%, por lo que la presencia de hombres en estos proyectos de calidad sigue siendo mayoritaria.

En el análisis no se percibe un patrón que indique que el número de mujeres haya aumentado con el tiempo. Su presencia no parece depender de este parámetro. Sin embargo, es importante recalcar, como se ha puntualizado en la metodología, que este análisis puede estar sesgado porque no se ha tenido en cuenta el número de mujeres en los cuatro proyectos transnacionales en los que han contribuido distintas organizaciones de todo el mundo y que suponen una parte importante de la muestra. Además, en esta categoría los equipos son más numerosos: se han identificado 112 periodistas involucrados en los proyectos ganadores, frente a 41 y 49 en los equipos de visualización y mejor app, respectivamente. El peso de estos proyectos recalca la importancia del periodismo colaborativo y transnacional en las investigaciones con datos.

5. DISCUSIÓN Y CONCLUSIONES

Esta investigación ha recopilado información sobre los proyectos de datos de primer nivel para discernir sobre su pervivencia, transparencia y paridad. El estudio, que ha tenido en cuenta una muestra de 35 proyectos premiados en un rango de siete años, refleja el buen estado del periodismo de datos en los últimos años. Del total de proyectos analizados, solo tres no están disponibles. Por su fecha de publicación (2013, 2015 y 2020), no se encuentra correlación entre el año de lanzamiento con su mejor o peor nivel de conservación. La variable del tiempo tampoco influye en el número de mujeres involucradas: no ha aumentado su presencia en los equipos de desarrollo analizados. Sin embargo, la transparencia de las metodologías sí ha mejorado en las últimas ediciones: desde 2017 empiezan a ser constante en las aplicaciones y en las investigaciones de datos.

A pesar de los buenos resultados obtenidos en el análisis de la pervivencia (PI1) es importante buscar alternativas que garanticen que los proyectos de datos sobrevivan al paso del tiempo. Además, el estudio plantea dudas sobre la pervivencia parcial de las piezas, pues se han encontrado proyectos en los que alguno de los apartados ya no funciona. Ocurría así con un vídeo alojado en YouTube del reportaje *Spies in the Skies*, publicado por BuzzFeed y reconocido en la categoría de mejor visualización en 2016. Como solución a esta problemática, Meredith Broussard propone crear distintas versiones estáticas de un proyecto. Ella sugiere transformar las páginas web en sitios dinámicos que puedan convertirse en un conjunto de páginas HTML estáticas. Éstas, al ser archivos de texto planos, tienen más posibilidades

de ser legibles en un futuro. “Las imágenes asociadas a las páginas web podrían representarse como TIFF, que se ha convertido en un formato muy popular para las imágenes archivadas” (Broussard, 2015a). En este caso, se reconoce una posible línea de investigación en la que profundizar sobre la conservación de cada uno de los elementos que forman los proyectos de datos y en soluciones para preservarlos.

En cuanto a la transparencia de los proyectos (PI2), el estudio refleja que es mayor y se tiene más en cuenta en las investigaciones de datos; es en esta categoría en la que las metodologías son más comunes. En el apartado de mejor aplicación, se observa una tendencia al alza con el paso del tiempo, pues todas a partir de 2017 empezaron a incluirla. Por lo tanto, se intuye que la concienciación sobre la importancia de esta práctica va en aumento. El uso de metodologías contribuye a que aumente la precisión de los estudios pues minimizan los riesgos de que los periodistas ofrezcan una información sesgada al compartir el por qué de utilizar ciertas fuentes. A pesar de sus puntos positivos, esta práctica todavía no es habitual en las visualizaciones de datos ya que solo dos proyectos de nuestro estudio las incorporaban. Esta carencia tiene un impacto negativo en la posibilidad de que las visualizaciones sean reproducidas por otros medios y que los lectores puedan acceder a la información sobre los datos utilizados (Marín y Arias, 2021)

El estudio de la paridad en los equipos de desarrollo de estos contenidos premiados muestra que la presencia de mujeres todavía no es igualitaria (PI3). Esto coincide con investigaciones relacionadas con la materia, tales como la encuesta de *Global Data Journalism Survey* de 2017, en la que se describía el estado del periodismo de datos a nivel global. Del total de periodistas que participaron en este estudio, 181, el 57,5% fueron hombres y el 42,5% mujeres (Heravi y Lorenz, 2020). Ocurre también así en la encuesta realizada por DataJournalism.com, *The State of Data Journalism Survey*, lanzada cuatro años más tarde, en 2021. En este caso, el paso del tiempo tampoco ha supuesto una mayor representación femenina: de una muestra de 1.274 participantes, el 59% eran hombres con respecto a un 39% de mujeres (DataJournalism.com, 2022). En la misma línea, el informe de la Data Visualization Society de 2021, que se centraba en mapear el panorama de los expertos en visualización de datos, fue contestado por un total de 1.739 personas. En este grupo, el 55.9% se identificaba como hombre frente al 43,1% de mujeres (DVS, 2022). Estas cifras coinciden con la representación de este estudio en el que las mujeres no superaban el 50% en ninguna de las disciplinas premiadas, siendo la categoría de mejor visualización en la que estaban menos representadas (31,07%).

Este trabajo presenta ciertas limitaciones. Los resultados obtenidos están condicionados por la elección de la muestra. Como se ha puntualizado en la metodología, se han dejado fuera del análisis varios proyectos por la imposibilidad de acceder a ellos o por la dificultad de evaluar sus características, un factor que limita la generalización de los resultados. También sería conveniente incorporar metodologías cualitativas (entrevistas,

grupos de discusión, cuestionarios...) que permitan profundizar en el trabajo de los periodistas de datos y conocer mejor sus prácticas profesionales y condicionantes en la producción de proyectos.

Además, es importante profundizar en las líneas de investigación abiertas. Por ejemplo, sería conveniente continuar indagando en las cuestiones de igualdad que afectan a estos equipos para entender cómo las redacciones se enfrentan al reto de construir un equipo inclusivo de periodistas de datos que no sea reflejo de un problema estructural más amplio (Howard, 2014). Por otro lado, en futuros estudios sería útil hacer hincapié en los roles ejercidos por cada uno de los profesionales de los equipos de datos para averiguar si hay diferencias entre las funciones que desempeñan hombres y mujeres. Este estudio se ha limitado a examinar cómo difiere el porcentaje entre los hombres y las mujeres en los equipos, pero la información se puede ampliar explorando cuál es la situación profesional de estas mujeres y si desempeñan funciones equiparables a las de sus compañeros o no, en cuanto al volumen de trabajo, salario, condiciones laborales, etc.

En este análisis se han identificados varios proyectos colaborativos impulsados por el ICIJ que han sido premiados en los últimos años. En estos equipos transnacionales, la transparencia es clave para que los datos puedan ser comunicados por redacciones de todo el mundo. Por eso, en próximas investigaciones sería apropiado explorar cómo estas colaboraciones periodísticas evolucionan y de qué forma se gestionan, qué herramientas se utilizan para coordinar estos equipos, y si la paridad es distinta según los países en los que se trabaja. Además, también sería conveniente extrapolar este estudio al resto de las categorías premiadas en estos certámenes, para incorporar conclusiones extensibles a otras disciplinas del periodismo de datos. Por otro lado, en próximos trabajos se podrían valorar proyectos publicados con mayor anterioridad y ampliar la muestra a reportajes que no hayan sido premiados, lo que contribuiría a una descripción más amplia de los parámetros analizados y proporcionaría resultados más fácilmente generalizables.

En definitiva, esta investigación contribuye a seguir profundizando en las características de los productos finales del periodismo de datos. Una de las fortalezas del estudio es que explora proyectos en todo el mundo, por lo que aporta una visión amplia de la disciplina que se ve respaldada por una selección sistemática de contenidos premiados internacionalmente. Además, se ha centrado en tres variables fundamentales para el desarrollo de la especialidad. Primero, se ha debatido sobre la preservación de los proyectos periodísticos digitales y la necesidad de conservarlos para salvaguardar la cobertura de la historia reciente. Después se ha incidido en la transparencia, ya que asegurar que las piezas sean lo más accesibles posibles fortalecerá la recepción positiva por parte de la audiencia y ayudará a que sean reproducidas por otros medios, contribuyendo a una mayor colaboración. Y, por último, se ha debatido sobre paridad, un factor clave para que se

minimicen los sesgos y se fomente que los equipos de periodismo de datos incluyan la representación de una sociedad justa y equitativa.

Es fundamental, por tanto, seguir estudiando los avances de esta disciplina que no deja de crecer dentro de las redacciones españolas. Ello permitirá fomentar su calidad en un entorno cada vez más exigente, en el que esta disciplina periodística debe responder a los criterios de igualdad y transparencia y conseguir que sus frutos perduren en el tiempo pese al constante avance tecnológico.

REFERENCIAS BIBLIOGRÁFICAS

- ANDERSON, C. W. (2018). *Apostles of Certainty: Data Journalism and the Politics of Doubt*. Oxford: Oxford University Press.
- ALEXANDER, M. (2016). A study of gender issues in data journalism.: <https://mospace.umsystem.edu/xmlui/bitstream/handle/10355/59161/projectreport.pdf?sequence=4>
- BOSS, K., y BROUSSARD, M. (2017). Challenges of archiving and preserving born-digital news applications. *IFLA Journal*, 43(2), 150–157. <https://doi.org/10.1177/0340035216686355>
- BRAVO, A. A. (2013). El periodismo de datos y la web semántica. CIC. *Cuadernos de Información y Comunicación*, 18, 99-116.
- BROUSSARD, M., y Boss, K. (2018). Saving Data Journalism. *Digital Journalism*, 6(9), 1206–1221. <https://doi.org/10.1080/21670811.2018.1505437>
- BROUSSARD, M. (2015a). Preserving news apps present huge challenges. *Newspaper Research Journal*, 36(3), 299–313. <https://doi.org/10.1177/0739532915600742>
- BROUSSARD, M. (2015b). Big Data in Practice. *Digital Journalism*, 4(2), 266–279. <https://doi.org/10.1080/21670811.2015.1074863>
- CASSIDY, W. (2008). Traditional in Different Degrees: The Professional Role Conceptions of Male and Female Newspaper Journalists. *Atlantic Journal of Communication*, 16(2), 105-117.
- CHAMBERS, L., BOUNEGRU, L., y GRAY, J. (2012). *The Data Journalism Handbook: How Journalists Can Use Data to Improve the News*. New York: O'Reilly Media.
- CHAPARRO-DOMÍNGUEZ M.R y DÍAZ-CAMPO J. (2021) Data Journalism and Ethics: Best Practices in the Winning Projects (DJA, OJA and Sigma Awards), *Journalism Practice*, online first. <https://doi.org/10.1080/17512786.2021.1981773>
- CÓRDOBA-CABÚS, A. (2020). Estándares de calidad en el periodismo de datos: fuentes, narrativas y visualizaciones en los Data Journalism Awards 2019. *Profesional de la información*, 29(3), e290328. <https://doi.org/10.3145/epi.2020.may.28>
- DUFFY, C. y FLYNN, K., (2021, 11 septiembre). *Some of the most iconic 9/11 news coverage is lost. Blame Adobe Flash*. CNN. <https://edition.cnn.com/2021/09/10/tech/digital-news-coverage-9-11/index.html>

- DataJournalism.com. (2021, 7 abril). Data-Driven Gold Standards: Award-Worthy Data Journalism. <https://datajournalism.com/read/handbook/two/situating-data-journalism/data-driven-gold-standards-what-the-field-values-as-award-worthy-data-journalism-and-how-journalism-co-evolves-with-the-datafication-of-society>
- DataJournalism.com. (2021, 1 septiembre). Archiving Data Journalism. <https://datajournalism.com/read/handbook/two/organising-data-journalism/archiving-data-journalism>
- DataJournalism.com. (2020, 5 febrero). Q&A with The Sigma Awards team - Newsletter. <https://datajournalism.com/read/newsletters/q-a-sigma-awards>
- Data Visualization: State of the Industry Survey 2021 i. (2021). Data Visualization Society. Recuperado 21 de marzo de 2022, de <https://www.DATAVISUALIZATIONSOCIETY.ORG/SURVEY>
- GARCÍA-AVILÉS, J. A. (2020). How news organizations face disruption in a networked environment: exploring the implementation of innovations in media labs. In: Peña-Fernández, S.; Meso-Ayerdi, K. y Larrondo-Ureta, A. (ed.) *Active audiences. Empowering citizens' discourse in the hybrid media system*. Madrid: McGraw Hill, pp. 55-66.
- GARCÍA-AVILÉS, J. A. (2021). Journalism as Usual? Managing Disruption in Virtual Newsrooms during the COVID-19 Crisis. *Digital Journalism*, 9(9), 1239–1260. <https://doi.org/10.1080/21670811.2021.1942112>
- GEN (Global Editors Network) (2015). *The Winners of the Data Journalism Awards 2015*. Mailchimp archive. <https://us3.campaign-archive.com/?u=9163cb87f87c5fb73d398ba6e&id=49e8254a6b&e>
- Hansen, K. A., y Paul, N. (2015). Newspaper Archives Reveal Major Gaps in Digital Age. *Newspaper Research Journal*, 36(3): 290–298. <https://doi.org/10.1177/0739532915600745>
- Hansen, K. A., y Paula, N. (2017). *Future-Proofing the News: Preserving the First Draft of History*. Lanham: Rowman & Littlefield Publishers.
- Heravi, B. R., & Lorenz, M. (2020). Data Journalism Practices Globally: Skills, Education, Opportunities, and Values. *Journalism and Media*, 1(1), 26–40. <https://doi.org/10.3390/journalmedia1010003>
- KALATZI, O., BRATSAS, C., y VEGLIS, A. (2018). The Principles, Features and Techniques of Data Journalism. *Studies in Media and Communication*, 6(2), 36-50. <https://doi.org/10.11114/smc.v6i2.3208>
- KRIPPENDORFF, K. (2018). *Content Analysis: An Introduction to Its Methodology* (4th ed.). Los Angeles, CA: Sage.
- LA-ROSA, L. y SANDOVAL-MARTÍN, T. (2016). La insuficiencia de la Ley de Transparencia para el ejercicio del Periodismo de datos en España. *Revista Latina de Comunicación Social*, 71, 1.208 -1.229 <http://www.revistalatinacs.org/071/paper/1142/62es.html> DOI: 10.4185/RLCS-2016-1142
- LOOSEN, W., REIMER, J., y DE SILVA-SCHMIDT, F. (2017). Data-driven reporting: An on-going (r)evolution? An analysis of projects nominated for the Data Journalism Awards 2013–2016. *Journalism*, 21(9), 1246–1263. <https://doi.org/10.1177/1464884917735691>

- LORD, C. L. (Ed.). (2016). *ElectionLand. Case Study 2016*. https://propublica.s3.amazonaws.com/assets/docs/electionland-case-study.pdf?_ga=2.3879155.259547948.1633186833-1958213824.1633186833
- MAIR, J., RADCLIFFE, D., y FELLE, T. (2015). *Data Journalism: Inside the Global Future*. Theschoolbook.com.
- MAIER, S. R. (2000). The Digital Watchdog's First Byte: Journalism's First Computer Analysis of Public Records. *American Journalism*, 17(4), 75–91. doi:10.1080/08821127.2000.10739272.
- MARÍN-SANCHÍZ, C. y ARIAS-ROBLES, F. (2021). El periodismo de datos es de ciencias: estrategias para incrementar la calidad del contenido a través del método científico. *Ciencia y Periodismo. Una es de Marte y otra es de Venus. XV Jornadas Internacionales de Innovación en Periodismo*. Elche: Editorial UMH, 133 – 144.
- ROGERS, S. (2014). “Data journalism is the new punk”. *British Journalism Review*, 25(2), 31-34. <https://doi.org/10.1177/0956474814538181>
- ROTHENBERG, J. (1999). “Ensuring the Longevity of Digital Information,” *clir.org*, February 22, 1999, <<http://www.clir.org/pubs/archives/ensuring.pdf>>
- SHOEMAKER, P. J., y REESE, S. D. (1996). *Mediating the message* (2nd ed.). White Plains, NY: Longman.
- STALPH, F. (2017). Classifying Data Journalism. *Journalism Practice*, 12(10), 1332–1350. <https://doi.org/10.1080/17512786.2017.1386583>
- State of Data Journalism 2021 Results*. (2022). DataJournalism.Com. Recuperado 21 de marzo de 2022, de <https://datajournalism.com/survey/2021/>
- STODDEN, V. y MIGUEZ, S. (2014). Best Practices for Computational Science: Software Infrastructure and Environments for Reproducible and Extensible Research. *Journal of Open Research Software*, 2(1), 1–6