

# **Evaluación de la transparencia en el periodismo de datos de referencia.**

## **Estudio de las historias publicadas entre 2018 y 2019**

*Evaluating the transparency in reference data journalism.  
Study of the stories published between 2018 and 2019*

*Avaliação da transparência no jornalismo de referência.  
Estudo de histórias publicadas entre 2018 e 2019*

**Alba Córdoba-Cabús**

*Personal Investigador en Formación  
(Universidad de Málaga)*

<https://orcid.org/0000-0002-3519-0583>

*España*

**Manuel García-Borrego**

*Personal Investigador Posdoctoral  
(Universidad de Málaga)*

<http://orcid.org/0000-0001-6207-8741>

*España*

**Fecha de recepción:** 25 de noviembre de 2020

**Fecha de revisión:** 6 de febrero de 2021

**Fecha de aceptación:** 24 de mayo de 2021

**Fecha de publicación:** 1 de julio de 2021

**Para citar este artículo:** Córdoba-Cabús, A. y García-Borrego, M. (2021). Evaluación de la transparencia en el periodismo de datos de referencia. Estudio de las historias publicadas entre 2018 y 2019, *Icono 14*, 19(2), 364-387. doi: 10.7195/ri14.v19i2.1648

## Resumen

*La aprobación de leyes que garantizan el acceso a la información y fomentan la transparencia de las instituciones propulsó el establecimiento del periodismo de datos como práctica periodística. Esta especialización surge en favor de la transparencia, por lo que incorporar indicadores que resten opacidad a la industria mediática se plantea como necesario para añadir credibilidad y calidad a la historia publicada. Esta investigación se propone examinar el nivel de transparencia en las mejores historias de periodismo de datos reconocidas a nivel mundial publicadas entre 2018 y 2019. Asimismo, se pretende comprobar si la tipología de las fuentes y de los datos inciden en la opacidad y clasificar las piezas en función del grado de transparencia. A través de un análisis descriptivo e inferencial de los proyectos candidatos a los Data Journalism Awards 2019 y a los Sigma Awards 2020 (n=80) se detecta un amplio margen de mejora en cuanto a la transparencia. Mientras que en la mayoría de las piezas se menciona directamente la fuente (91,4%), el acceso a la metodología (48,8%) y a los datos en sí (20%) resulta muy limitado. Del análisis de correspondencias múltiple se desprenden tres grupos de historias, buena parte de las cuales pertenecen al nivel más opaco. Entre otras constataciones, se concluye que el tipo de fuentes y de datos incorporados no influye en el nivel de transparencia.*

**Palabras clave:** *Transparencia; Periodismo de datos; Datos; Data Journalism Awards 2019; Sigma Awards 2020; Opacidad*

## Abstract

*The approval of laws that provided access to information and promote the transparency of institutions favored the establishment of data journalism as a journalistic practice. This specialization arises in favor of transparency, so incorporating indicators that reduce opacity of the media industry is necessary to add credibility and quality to the published story. This research aims to examine the level of transparency in the best data journalism stories published worldwide between 2018 and 2019. It also aims to check whether the typology of sources and data affects opacity and to classify the pieces according to the degree of transparency. Through a descriptive and inferential analysis of the candidate projects for the Data Journalism Awards 2019 and the Sigma Awards 2020 (n=80), a wide range of improvement in terms*

*of transparency is detected. While the vast majority of pieces mention the source directly (91.4%), access to the methodology (48.8%) and to the data (20%) is quite limited. The multiple correspondence analysis reveals three groups of stories, many of which belong to the highest level of opacity. Among other findings, it is concluded that the type of sources and data incorporated does not influence on the level of transparency.*

**Keywords:** *Transparency; Data Journalism; Data; Data Journalism Awards 2019; Sigma Awards 2020; Opacity*

## **Resumo**

*A adoção de leis que garantem o acesso à informação e promovem a transparência das instituições favoreceu o estabelecimento do jornalismo de dados como uma prática jornalística. Esta especialização surge a favor da transparência, pelo que é necessário incorporar indicadores que reduzam a opacidade da indústria dos media para acrescentar credibilidade e qualidade à história publicada. Esta investigação visa examinar o nível de transparência nas melhores histórias de jornalismo de dados globalmente reconhecidas, publicadas entre 2018 e 2019. Visa também verificar se a tipologia das fontes e dos dados afecta a opacidade e classificar as peças de acordo com o grau de transparência. Através de uma análise descritiva e inferencial dos projectos candidatos aos Data Journalism Awards 2019 e aos Sigma Awards 2020 (n=80), é detectada uma ampla margem de melhoria em termos de transparência. Embora a maioria das peças mencione directamente a fonte (91,4%), o acesso à metodologia (48,8%) e aos próprios dados (20%) é muito limitado. A análise de correspondência múltipla revelou três grupos de histórias, muitas das quais se encontravam ao nível mais opaco. Entre outras conclusões, conclui-se que o tipo de fontes e dados incorporados não influencia o nível de transparência.*

**Palavras chave:** *Transparência; Jornalismo de dados; Dados; Data Journalism Awards 2019; Sigma Awards 2020; Opacidade*

## 1. Introducción

La transparencia se entiende como un término vinculado con el hecho de que las empresas, los gobiernos y las instituciones pongan a disposición de los usuarios información real de interés público, respondiendo así al derecho a saber de la ciudadanía y permitiéndoles tomar decisiones con conocimiento de causa (Lizcano, 2012). La transparencia está ligada a la buena gestión de las organizaciones, que la han incorporado a sus estrategias con la intención de frenar el auge de la desconfianza de la opinión pública y con el firme convencimiento de que esta incide positivamente en su reputación y en la capacidad persuasiva de sus mensajes (García, 2014).

En el ejercicio del periodismo, la transparencia se considera un principio ético esencial. Su adhesión en 2014 al código ético de la Sociedad de Periodistas Profesionales planteó un nuevo paradigma al que los medios de comunicación debían enfrentar, entendiéndolo como una oportunidad para reforzar las relaciones de los periodistas con las instituciones y la ciudadanía (Kovach & Rosentiel, 2001; Deuze, 2003; Hayes, Singer & Ceppos, 2007; Karlsson, 2010; Manfredi, 2014; La-Rosa & Sandoval-Martín, 2016; Vos & Craft, 2017; Karlsson & Clerwall, 2018).

El concepto de transparencia en la práctica periodística, conocida como “transparencia mediática” (Campos-Domínguez & Redondo-García, 2015; Díez-Garrido & Campos-Domínguez, 2018), implica divulgar tantos detalles como sea posible sobre el proceso de producción de la noticia (Allen, 2008; Karlsson, 2010; Karlsson & Clerwall, 2018; Appelgren & Salaverría, 2018). Esto supone que el periodista ha de ser honesto con la audiencia y debe explicar cómo se recopiló la información, qué métodos se utilizaron para analizarla e incluso qué factores intervinieron en el proceso (Kovach & Rosentiel, 2001; Gehrke, 2020). Karlsson (2010) discernió entre dos estadios de transparencia: el revelador, aquel que explica la selección y la producción de la información; y el participativo, que invita a la audiencia a colaborar en el proceso de elaboración.

Pese a que diversos académicos atestiguan que la transparencia repercute en la confianza y ayuda a recuperar la credibilidad perdida (Weinberger, 2009; Singer, 2010; Philips, 2010; Vos & Craft, 2017; Anderson & Borges-Rey, 2019), otros consideran que su influencia como elemento regenerador es limitada (Karlsson, Clerwall

& Nord, 2017; Anderson, 2018; Karlsson, 2020). A pesar de las discrepancias suscitadas, resulta plausible pensar que la transparencia contribuye eficazmente, y más en un contexto como el actual, a combatir la desinformación.

### **1.1. El papel de la transparencia en el periodismo de datos**

El establecimiento de leyes que garantizan el acceso a la información y promueven la transparencia fomenta el surgimiento de proyectos basados en datos (Cortés del Álamo, Luengo Cruz & Elías, 2018). La instauración, en parte generalizada, de una cultura de datos abiertos, impulsó la eclosión de figuras profesionales relacionadas con la gestión de cantidades ingentes de información como el periodista de datos (Paraise & Darigal, 2012; Solop & Wonders, 2016). Esta práctica nace en favor de la transparencia (Beiler, Irmer & Breda, 2020) y permite, en cierta medida, alejar al periodismo de las posturas que lo califican como opaco (Plaisance, 2000). Tanto es así que diversos autores como Martinisi (2013), Hammond (2017), Jamil (2019) y Gehrke (2020) asumen que la transparencia es una de las características definitorias del periodismo de datos, aunque la forma de implementarla difiera en función de la empresa mediática, el país y el sistema de medios (Appelgren & Salaverría, 2018).

Meyer (1991) expresó la necesidad de introducir detalles sobre el origen de la información y la metodología empleada en el periodismo basado en datos. En base a estas declaraciones, la transparencia comenzó a denominarse “nueva objetividad”, entendiéndose que este término era la evolución de la tradicional objetividad. Weinberger (2009) aclaró que esta, la transparencia, predominaba en la era digital, mientras que la objetividad lo hacía cuando prevalecía el papel. Posteriormente, Allen (2008) secundó la postura de Meyer (1991) y expresó la necesidad de aplicar los ideales de transparencia en aras de intensificar la legitimidad de las organizaciones periodísticas. Si bien, como Appelgren y Salaverría (2018) sugirieron, para que este principio genere impacto en la audiencia debe manifestarse en el contenido informativo.

Los datos suscitan neutralidad (Porter, 1995; Tal & Wansink, 2016; Kennedy, Weber & Engebretsen, 2020), pero su manejo conlleva innumerables decisiones por parte del periodista. Ante este control de los puntos de vista, ofrecer la oportunidad de reutilizar las cifras y conocer la metodología del trabajo incrementaría la

## ARTÍCULOS DE INVESTIGACIÓN

calidad de las historias y permitiría reforzar la función de perro guardián de los medios de comunicación (Philips, 2010; Alexander & Vetere, 2011; Paraise & Darigal, 2012; Howard, 2014; Coddington, 2015; Solop & Wonders, 2016; Felle, 2016; Hammond, 2017; Horky & Pelka, 2017; Anderson & Borges-Rey, 2019).

Fink y Anderson (2015) y Kennedy, Weber y Engebretsen (2020) expusieron que para los profesionales de la información la transparencia está instaurada en el ejercicio del periodismo de datos. Sin embargo, Solop y Wonders (2016) y Tandoc y Oh (2017) insinuaron que el escaso reconocimiento y valoración del periodismo de datos se debía a la opacidad en los procesos. Hace ya más de una década, Hayes, Singer y Ceppos (2007) reconocieron que la transparencia era uno de los retos contemporáneos de la profesión periodística, mientras Porlezza y Splendore (2019), años más tarde, mantienen que todavía lo sigue siendo y atribuyen su escasa presencia al temor de los propios periodistas a que imiten o plagien su trabajo.

El número de investigaciones centradas en evaluar la transparencia en los medios de comunicación resulta insuficiente y aquellas que la estudian exhiben resultados dispares sin alcanzar un consenso sobre tendencias o patrones establecidos en las redacciones. En el último lustro, se detecta un predominio mayor de los estudios que aluden a la opacidad de las informaciones frente a las que reflejan transparencia.

Karlsson (2010), a través de un estudio de las principales versiones *online* de diarios estadounidenses, británicos y suecos, evidenció cómo la transparencia empezaba a manifestarse, aunque el grado y el modo de implantación difería según la procedencia del medio. Por un lado, Paraise y Darigal (2012), Tandoc y Oh (2017) y Loosen, Reimer y Schmidt (2017) localizaron atisbos de transparencia en más de la mitad de las piezas analizadas, bien a través de la publicación de los datos en bruto, bien mediante la incorporación de detalles metodológicos. Por otro lado, Stalph (2017), Lowrey y Hou (2018), Young, Hermida y Fulda (2018), Zhang y Feng (2019) y Zamith (2019) constataron la limitada manifestación de indicadores de transparencia en los trabajos objeto de estudio, llegando incluso a calificarlos de opacos. A pesar de las divergencias detectadas, las investigaciones coinciden en revelar cómo las historias de periodismo de datos están supeditadas a la información pública procedente de instituciones, gobiernos y oficinas públicas, considerándolo esto una limitación para la propia transparencia.

En definitiva, en aras de incrementar la confiabilidad de la información periodística se precisa de publicaciones abiertas que posibiliten una aproximación al proceso de producción y a los datos con los que se confecciona la noticia. Dada la importancia de integrar estos estándares en los artículos basados en datos, en este trabajo se proponen cuatro objetivos:

1. Determinar la presencia de elementos de transparencia en las historias de periodismo de datos de referencia a nivel mundial publicadas entre 2018 y 2019, teniendo en cuenta para ello las fuentes, las cifras y los análisis efectuados.
2. Establecer las propiedades de los indicadores de transparencia en los proyectos nominados a los *Data Journalism Awards 2019* y los *Sigma Data Journalism Awards 2020*.
3. Comprobar si la tipología de las fuentes y de los datos inciden en el nivel de transparencia de la información.
4. Clasificar, a partir del análisis de los nominados a los *Data Journalism Awards 2019* y los *Sigma Awards 2020*, en grupos las historias de periodismo de datos en función del grado de transparencia.

## 2. Material y métodos

### 2.1. Muestra

Para delimitar cuáles eran las piezas consideradas de calidad en esta especialización durante este periodo, el estudio se centró en el análisis de los nominados a los *Data Journalism Awards 2019* –trabajos publicados entre el 26 de marzo de 2018 y el 7 de abril de 2019– y los *Sigma Data Journalism Awards 2020* –proyectos difundidos en 2019–. Ambos galardones, aunque con distinta nomenclatura, emergen con un propósito similar: premiar las mejores historias de periodismo de datos a nivel mundial y agasajar el trabajo de los periodistas. Los *Data Journalism Awards*, convocados por la Global Editors Network, celebraron su última edición en 2019 y los *Sigma Data Journalism Awards*, fundados por Aron Pilhofer y Regina Chua y apadrinados por Google News Initiative, tomaron el testigo. Este análisis permitirá conocer el estado actual de la transparencia en el periodismo de datos sin obstáculos temáticos ni geográficos.

## ARTÍCULOS DE INVESTIGACIÓN

Para el examen se identificaron y suprimieron aquellos trabajos duplicados, con nominación en 2019 y 2020, para evitar reiteraciones. Asimismo, se extrajeron de la muestra las nominaciones a perfiles profesionales o equipos completos, aplicaciones y páginas webs de datos abiertos. Las categorías susceptibles de análisis fueron: innovación, periodismo de datos en menos de 36 horas, visualización, investigación y premio del público en la convocatoria de 2019 y mejor historia de periodismo de datos, innovación y mejor visualización en la de 2020. La muestra final quedó compuesta por 80 proyectos (40 piezas de cada convocatoria).

### 2.2. Instrumento

La evaluación de la transparencia ha experimentado cambios y los estudios centrados en examinar si determinadas publicaciones cumplían esta exigencia no coinciden en evaluar unas categorías comunes. Hayes, Singer y Ceppos (2007) evaluaron la transparencia mediante el análisis de las fuentes y la apertura de los datos, estándares replicados posteriormente por Karlsson (2010), Karlsson y Clerwall (2018) y Zhang y Feng (2019) en sus investigaciones. Por su parte, Paraise y Darigal (2012), Loosen, Reimer y Schmidt (2017), Young, Hermida y Fulda (2018) y Zamith (2019) apostaron por registrar el acceso a las cifras en bruto y la aparición de detalles metodológicos, mientras que Tandoc y Oh (2017), Stalph (2017) y Lowrey y Hou (2018) se ciñeron a evaluar la transparencia en función de la ausencia o presencia de la opción de descarga del *dataset*.

En esta investigación se aplicó un análisis de contenido cuantitativo con carácter descriptivo e inferencial a las mejores historias de periodismo de datos publicadas entre 2018 y 2019 a nivel mundial. El establecimiento de las variables se fundamentó en las consideraciones plasmadas en la literatura científica y se anexionaron otras *ad hoc* para dar respuesta a los objetivos planteados en esta investigación. El libro de códigos propuesto permitía inspeccionar la transparencia en tres niveles: la procedencia de la información, las características numéricas y los cálculos o procesos llevados a cabo (Tabla 1).



Dimensión	Variables	Categorías
<b>Fuentes</b>	Número	
	Tipo	Pública
		Privada
		Otras organizaciones
		Fuentes propias
		Filtraciones
		Otro
	Atribución	Directa
		Con reserva
<i>Off the record</i>		
<b>Datos</b>	Grado de transparencia	No mención fuente – No metodología
		Sí mención fuente – No metodología
		No mención fuente – Sí metodología
		Sí mención fuente – Sí metodología
	Acceso a los datos	No
		Sí
	Tipo	Geodatos
		Sociodemográficos
		Personales
		Medidos
		Médicos
		Encuestas
		Financieros
		Metadatos
Otros		
<b>Procesos estadísticos</b>	Metodología	No
		Sí
	Forma de incorporación	Dentro del propio texto
		En un apartado específico

**Tabla 1:** Variables susceptibles de análisis. Fuente: elaboración propia.

## ARTÍCULOS DE INVESTIGACIÓN

La primera dimensión, las fuentes, se evaluó en base al número, el tipo –gobierno u oficina pública, corporaciones privadas, otras organizaciones, filtraciones y fuentes propias– y la atribución –distinguiendo entre atribución directa, con reserva y *off the record*– (Hayes, Singer & Ceppos, 2007; Karlsson, 2010; Karlsson & Clerwall, 2018; Zhang & Feng, 2019).

En el segundo nivel, los datos, se determinó el grado de transparencia de estos –en función de la mención al origen de la información y las puntualizaciones sobre el *dataset*–, si se permitía el acceso a los datos en bruto y el tipo (Hayes, Singer & Ceppos, 2007; Karlsson, 2010; Paraise & Darigal, 2012; Loosen, Reimer & Schmidt, 2017; Stalph, 2017; Tandoc & Oh, 2017; Lowrey & Hou, 2018; Young, Hermida & Fulda, 2018; Karlsson & Clerwall, 2018; Zamith, 2019; Zhang & Feng, 2019). En este punto se podía discernir entre:

- Geodatos. Especificaciones sobre ubicaciones geográficas. Se incorporan cuando se emplea un mapa como visualización.
- Sociodemográficos. Características relativas a un segmento poblacional. Por ejemplo: religión, ingresos, nivel educativo, etc.
- Personales. Detalles concretos de un sujeto. Por ejemplo: biografía, peso, estatura, ingresos, etc.
- Medidos. Valores recopilados mediante herramientas de medición o sensores.
- Médicos. Información sanitaria. Por ejemplo: altas, ingresos, capacidad de las UCI, disponibilidad de material, etc.
- Encuestas. Pormenores sobre la opinión pública o puntos de vistas.
- Financieros. Apreciaciones económicas.
- Metadatos. Descriptores de los datos. Por ejemplo: datos de correo electrónico, sobre el uso de las redes sociales, de telecomunicaciones, etc.
- Otros. Cualquiera que no pueda suscribirse a las categorías anteriores.

En el tercer escalafón, referente a los procesos estadísticos, se valoró la incorporación de aclaraciones metodológicas sobre el trabajo con cifras y la forma en que estas se plasmaban en la publicación –en un apartado específico o en el cuerpo

de la noticia– (Paraise & Darigal, 2012; Loosen, Reimer & Schmidt, 2017; Young, Hermida & Fulda, 2018; Zamith, 2019).

En esta investigación se instauran los siguientes requisitos fundamentales para la consecución de un grado óptimo de transparencia: mencionar directamente a la fuente o, en su defecto, al entorno del que procede la información, aportar detalles sobre las características del *dataset*, permitir el acceso total a los datos y añadir pinceladas metodológicas sobre el trabajo con cifras.

### 2.3. Procedimiento

A partir de los proyectos de la muestra se construyó una matriz de datos en el software estadístico SPSS, con el que se realizaron los análisis descriptivos e inferenciales para determinar la incidencia del tipo de fuentes y de datos en el grado de transparencia de los nominados a los *Data Journalism Awards* 2019 y los *Sigma Data Journalism Awards* 2020.

Para detectar la asociación entre variables nominales se calculó el estadístico de contraste chi-cuadrado o, en su defecto, la corrección de continuidad de Yates –en caso de contar con un grado de libertad–. Si el *test* de independencia resultaba significativo ( $p \leq 0.05$ ), se examinaban los residuos tipificados corregidos con el propósito de definir la dirección de la asociación y el tamaño del efecto provocado.

Con la intención de identificar semejanzas entre piezas de periodismo de datos y establecer grupos en función de la transparencia se optó por el análisis factorial de correspondencias múltiple, que permite entrever las relaciones entre un conjunto de variables nominales y representar en un espacio reducido las dimensiones de nuestro conjunto de datos. Se empleó un análisis biplot para explorar las relaciones existentes entre las diferentes categorías de las variables y los trabajos seleccionados. En este caso, el estudio se efectuó con el software libre R, debido a la diversidad y potencialidad de las visualizaciones que genera.

### 3. Resultados

A continuación, se exponen los hallazgos obtenidos de forma pormenorizada haciendo hincapié en las relaciones encontradas entre las variables de estudio.

#### 3.1. Tipología de fuentes y atribución

Los proyectos examinados integran una media de tres fuentes ( $M=2,64$ ,  $DT=2,11$ ), sin embargo, lo más habitual es la incorporación de información procedente de un único origen (moda=1; 29,6%). Se localizan 18 trabajos en los que o bien no se especifica la cuantía exacta de fuentes empleadas, o bien la cantidad era muy elevada –en algunos casos superior a 20–, por lo que se optó por clasificarlos como valores perdidos y aislarlos de la muestra para no desvirtuar los estadísticos de tendencia central.

Los nominados a los premios de periodismo de datos se sustentan de fuentes públicas, evidenciándose una fuerte dependencia de la información procedente de la administración y de oficinas oficiales (60%) como el National Center for Education Statistics Education Demographic and Geographic Estimates de Estados Unidos o la Office for National Statistics de Reino Unido. Las piezas exponen un 36,3% de cifras procedentes de organizaciones como centros de investigación, universidades u ONG, seguido de cerca de la información reunida por los propios medios o entidades que presentan su candidatura (35%) –aquí se incluyen la realización de encuestas, *scrapping* en web o documentos, así como análisis de redes sociales–. En menor medida se hallan datos compilados por corporaciones privadas (17,5%), valores extraídos de emisiones televisivas (albergados en la categoría “otros” con un 11,3%) o filtraciones (3,8%). Se detecta un proyecto, titulado “Key events from past four months of Hong Kong’s anti-government protests” (*South China Morning Post*), que no facilita ni el origen exacto de la información ni una aproximación al contexto (1,3%).

En términos generales, los candidatos de ambas ediciones mencionan el origen concreto o el entorno del que proceden las cifras. Solo en el 8,6% de los casos se reserva la identidad de la fuente o no se menciona, mientras que en el resto se ci-

tan con exactitud (91,4%). La atribución reservada se corresponde, en su mayoría, con publicaciones derivadas de filtraciones, como ocurre en “Driver’s notebooks exposed Argentina’s greatest corruption scandal ever: ten years and millions of cash bribes in bags” de *La Nación* (Argentina), en “Copy, paste, legislate” de *Usa Today*, *The Arizona Republic* y Center for Public Integrity y en “How life has changed for young people your age” de ABC News.

Tipo de fuentes	Atribución		Total	p-valor
	Directa	Reservada		
<b>No hay fuente</b>	-	-	1,3%	-
<b>F. Públicas</b>	95,8%	4,2%	60,0%	p=0,118
<b>F. Privadas</b>	100%	0,0%	17,5%	p=0,969
<b>Otras org.</b>	96,6%	3,4%	36,3%	p=0,472
<b>F. Propias</b>	96,4%	3,6%	35,0%	p=1,000
<b>Filtración</b>	0,0%	100%	3,8%	p=0,000*
<b>Otras</b>	88,9%	11,1%	11,3%	p=1,000

**Tabla 2:** Asociación entre el tipo de fuente y la atribución en los proyectos nominados.

Las informaciones que emplean la atribución reservada emanan, principalmente, de filtraciones (100%), de instituciones u oficinas públicas (4,2%), de cifras recopiladas por el propio equipo (3,6%) y de otras organizaciones (3,4%). Si bien, la corrección de continuidad de Yates vislumbra que el tipo de fuente empleada solo incide en la atribución cuando se trata de valores filtrados (Tabla 2). En este sentido la relación resulta estrecha (TE=1, efecto alto): difundir información de este estilo conlleva no referenciar directamente a la fuente.

### 3.2. Tipos de datos, transparencia y acceso

El tipo de dato más frecuente en los nominados a los premios de periodismo de datos es el sociodemográfico (42,5%), lo que responde a una elevada abundancia de cifras relacionadas con las características de distintos grupos de la población (ingresos, educación, situación laboral...). Los valores referentes a la localización –también conocidos como geodatos– se emplean de manera similar a los anteriores (36,3%), seguidos de cerca por los recopilados por sensores o herramientas de me-

## ARTÍCULOS DE INVESTIGACIÓN

dición (30%) y los metadatos (20%). En menor medida se utilizan los financieros (18,8%), los indicadores personales (16,3%), los resultados procedentes de encuestas (15%) y los médicos (3,8%).

En general, los trabajos nominados en la edición de 2019 y 2020 se centran en identificar las fuentes de manera apropiada –directa o indirectamente–, pero no añaden información sobre el conjunto numérico (61,2%), de manera que impiden descubrir cómo se estructura la información y cuáles son sus características. Esto ocurre en piezas como “Indonesia Plane Crash” de Reuters, “Every time Ford and Kavanaugh dodged a question, in one chart” de Vox o “Made in France” de Disclose.

Solo el 37,5% de los proyectos responde al ideal de transparencia en este aspecto, ya que incorpora pormenores tanto de la procedencia de las cifras como de la base de datos. Ejemplo de lo indicado son “O que revela uma análise das emoções dos candidatos durante o debate” de *O Estado de S. Paulo*, “The Invisible Crime” de *The Age*, *Sidney Morning*, *Brisbane Times* y *WA Today* y “The Quiet Rooms” de *Chicago Tribune* y *Propublica Illinois*. En el lado opuesto se encuentran las piezas en las que la fuente y los detalles sobre el *dataset* están ausentes (1,3%). En esta ocasión solo se localiza un artículo totalmente opaco y coincide con el ya mencionado “Key events from past four months of Hong Kong’s anti-government protests” (*South China Morning Post*). No se detecta ninguna pieza que no indique la fuente e incluya especificaciones sobre el conjunto de datos.

Se aprecia que los trabajos nominados más transparentes, en tanto que citan la fuente y adhieren información sobre el conjunto de datos, surgen, principalmente, a partir de filtraciones (66,7%) y valores compilados por el propio equipo o periodista (50%). Se infiere que, al tratarse de piezas en las que la atribución suele ser reservada –filtraciones– y las cifras las reúne o evalúa el periodista, resulta necesario aportar detalles sobre el *dataset* y aproximarse a la fuente para dotar de rigor a la publicación. Pese a identificarse diferencias en este sentido, del estadístico de contraste chi-cuadrado se desprende que, a un nivel de confianza del 95%, estas no resultan significativas, por lo que el manejo de un tipo de fuente o dato específico no influye en el nivel de transparencia alcanzado. Si bien, como se observa en la Tabla 3, el déficit de fuentes sí implica la opacidad de la pieza en cuestión.

Tipología	Transparencia datos <sup>1</sup>				p-valor
	0	1	2	3	
Tipo de fuentes					
<b>No hay fuente</b>	100%	-	-	-	p=0,000*
<b>F. Públicas</b>	-	64,6%	-	35,4%	p=0,395
<b>F. Privadas</b>	-	64,3%	-	35,7%	p=0,882
<b>Otras org.</b>	-	69,0%	-	31,0%	p=0,468
<b>F. Propias</b>	-	50,0%	-	50,0%	p=0,200
<b>Filtración</b>	-	33,3%	-	66,7%	p=0,565
<b>Otras</b>	-	66,7%	-	33,3%	p=0,895

**Tabla 3:** Incidencia del tipo de fuente en la transparencia de la información.

Los actores encargados de la producción de las publicaciones se muestran reticentes a facilitar la descarga de las cifras en la propia web del medio u organización (80%). Solo en el 20% se ofrece esta opción, entre los que se localizan “Dying homeless: counting the deaths of homeless people across the UK”, publicado por The Bureau of Investigative Journalism, “The Opioid Files” del *Washington Post* o “See How the World’s Most Polluted Air Compares With Your City’s” del *The New York Times*. En estos casos, la descarga se puede efectuar del total de la información numérica. Los proyectos que contienen datos derivados de filtraciones son los que permiten un acceso mayor a las cifras (66,7%). Por su parte, los que dependen de fuentes privadas (85,7%) y públicas (83,3%) e introducen geodatos (93,1%) son los más opacos. Al igual que ocurría en la variable anterior, el estadístico de contraste refleja que la incorporación de un tipo de dato o fuente concreta no incide en la apertura numérica.

### 3.3. Detalles metodológicos

Al examinar la presencia de páginas extra, *pop-up*, recuadros o párrafos en los que se explican los pormenores del análisis con cifras se vislumbra que más de la mitad de los trabajos no incorpora este tipo de especificaciones (51,2%). Esto evidencia la carencia de detalles metodológicos en los trabajos de periodismo de datos de calidad (ausencia de información sobre la recopilación, limpieza de los datos, análisis desarrollado, etc.). Solo el 48,8% de los nominados a estos premios incluye este conte-

## ARTÍCULOS DE INVESTIGACIÓN

nido, entre los que destacan piezas como “How top health websites are sharing sensitive data with advertisers” del *Financial Times* o “How to profit in space: a visual guide” del *The Wall Street Journal*. Del análisis pormenorizado de los resultados se percibe mayor predilección por incorporar las aclaraciones de este estilo en apartados específicos (61,4%) que por añadirlos en el propio cuerpo de la noticia (18,8%).

Tipología	Metodología <sup>2</sup>			p-valor
	0	1	2	
Tipo de fuentes				
<b>Geodato</b>	51,7%	20,7%	33,3%	p=0,913
<b>Sociodemográfico</b>	44,1%	17,6%	38,2%	p=0,375
<b>Personal</b>	46,2%	30,8%	23,1%	p=0,467
<b>Médicos</b>	33,3%	-	66,7%	p=0,337
<b>Medidos</b>	45,8%	29,2%	25,0%	p=0,291
<b>Encuestas</b>	41,7%	25,0%	33,3%	p=0,741
<b>Financieros</b>	33,3%	6,7%	60,0%	p=0,077
<b>Metadatos</b>	25,0%	37,5%	37,5%	p=0,033*
<b>Otros</b>	66,7%	14,8%	18,5%	p=0,134

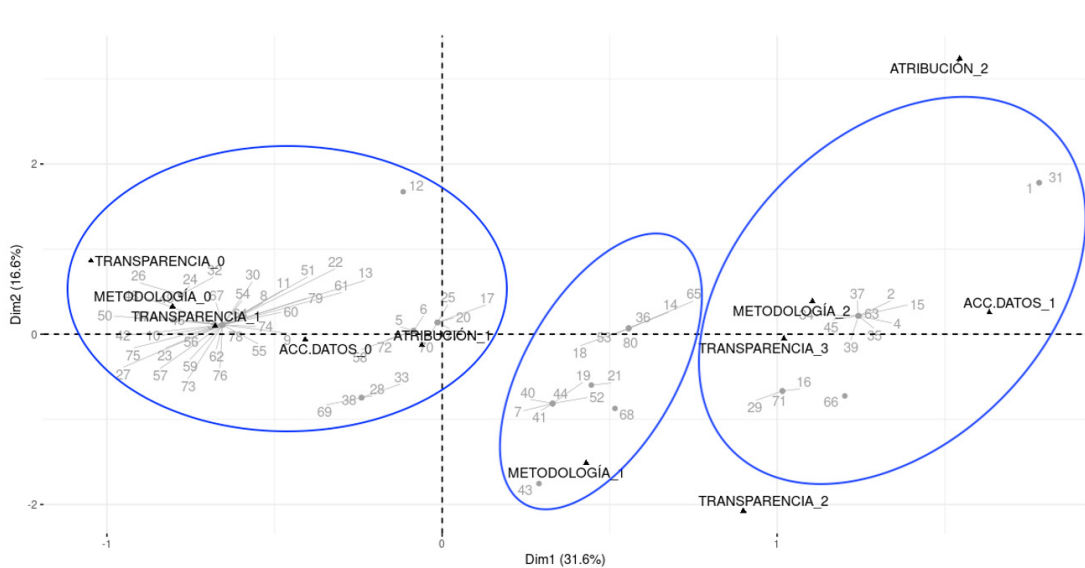
**Tabla 4:** Incidencia del tipo de dato en la incorporación de metodología.

En las piezas nominadas que integran cifras referentes a emisiones televisivas (otros, 66,7%), geodatos (51,7%) e información numérica personal (46,2%) es menos común la presencia de una metodología detallada. Cuando aparecen estos tipos de datos rara vez se describe cómo se ha desarrollado el trabajo. En el lado opuesto, las más transparentes en cuanto a la incorporación de una metodología se refiere, se localizan las publicaciones con metadatos (75%, el 37,5% la incluye dentro del propio texto y el 37,5% restante en un apartado específico). A excepción de estos últimos, el estadístico de contraste chi-cuadrado revela la inexistente asociación entre los distintos tipos de datos y la inclusión de metodología (Tabla 4). En este sentido, añadir valores relativos a las telecomunicaciones, sobre aplicaciones, correos electrónicos o uso de las redes sociales –conocidos como metadatos– está intrínsecamente relacionado con la aparición de especificaciones metodológicas en el texto (TE= 0,292, efecto moderado). Por tanto, en base a los análisis realizados en los distintos apartados, se podría afirmar que el tipo de fuente y de dato no incide en la transparencia de la información.



### 3.4. Tipos de trabajos en función de la transparencia

Una vez comprobado que la incorporación de un tipo de fuente o dato específico no influye en el grado de transparencia, se realizó un análisis factorial de correspondencias múltiple con las siguientes variables: atribución, transparencia de los datos, acceso a las cifras y metodología. La Figura 1 plasma la componente 1 (Dim1) con una inercia del 31,6%, mientras que la componente 2 (Dim2) posee un 16,6%. Si se observa el eje de la X, se aprecia una representación de las variables en orden creciente en función de la transparencia (de izquierda a derecha).



**Figura 1:** Análisis de correspondencias múltiple<sup>3</sup>.

A partir del análisis de correspondencias múltiple se puede esbozar una diferenciación de tres núcleos de proyectos en función del grado de transparencia:

- Grado alto. Engloba trabajos que se asemejan al ideal de transparencia marcado. Mencionan la fuente o, en su defecto, el entorno de la información, proporcionan detalles sobre el conjunto de datos, especifican la metodología –preferentemente en el texto de la noticia– y permiten el acceso a los datos.
- Grado medio. En este nivel intermedio se encuentran aquellas piezas que presentan leves atisbos de transparencia, pero no cumplen con el total de los

## ARTÍCULOS DE INVESTIGACIÓN

indicadores. La principal diferencia con el grado anterior radica en el acceso a las cifras, ya que aquí no se permite su descarga. A su vez, este grupo incluye la metodología, pero opta por integrarla en un apartado específico.

- Grado bajo. Citan la fuente o se aproximan a ella, pero carecen de especificaciones sobre las características del *dataset* y los análisis realizados y no facilitan la descarga de la información. Aquí también se ubica el proyecto que no menciona el origen de las cifras. La mayor parte de los proyectos nominados a estos premios se encuadran en este núcleo.

## 4. Discusión y conclusiones

La presente investigación pone de manifiesto la importancia de la transparencia y evidencia el aún holgado margen de mejora en las historias de periodismo de datos de referencia, aquellas nominadas a los *Data Journalism Awards 2019* y a los *Sigma Awards 2020*.

01 y 02. A la luz de los resultados obtenidos, se observa un exiguo uso de los indicadores de transparencia, aunque este déficit no se aprecia en la misma medida en las tres perspectivas examinadas. Mientras que en la mayoría de los trabajos se citan directamente las fuentes, el acceso a detalles metodológicos y a los datos en sí resulta muy limitado. Estos hallazgos apuntalan lo mencionado por autores como Stalph (2017), Lowrey y Hou (2018), Young, Hermida y Fulda (2018) Zhang y Feng (2019) y Zamith (2019), quienes destacan la tímida implantación de la transparencia en los proyectos basados en datos. Se podría decir que la dimensión vinculada a las cifras es la más opaca, quizás explicado por la falta de recursos, tiempo y conocimientos por parte de los periodistas para facilitar dicha apertura o, como señalaban Porlezza y Splendore (2019), por el temor a que repliquen su trabajo. Como se ha recogido en este estudio, siempre que se permite la descarga puede efectuarse del total de las cifras y en aquellos casos en los que se incluye la metodología aparece en un apartado específico para ello.

03. Los estadísticos de contraste chi-cuadrado y la corrección de continuidad de Yates reflejan que, por lo general, la tipología de las fuentes o de los datos incorporados no inciden en la transparencia de la información. Sin embargo, se registran

algunas relaciones entre, por ejemplo, el uso de datos procedentes de filtraciones y el empleo de la atribución reservada y el manejo de metadatos y la aparición de metodología. Esto atenuaría la consideración que alude a la dependencia de fuentes gubernamentales como el principal motivo de la opacidad de las historias de periodismo de datos (Paraise & Darigal, 2012; Tandoc & Oh, 2017; Loosen, Reimer & Schmidt, 2017; Stalphy, 2017; Lowrey & Hou, 2018; Young, Hermida & Fulda, 2018; Zhang & Feng, 2019; Zamith, 2019) y reforzaría la postura sobre la falta de recursos, tiempo y conocimientos establecida anteriormente.

04. A partir del análisis factorial de correspondencias múltiple se distinguieron tres grupos de proyectos basados en el grado de transparencia (alto, medio y bajo). Se aprecia que la mayoría de los trabajos examinados se encuadran en el nivel inferior, ya que priorizan la mención de las fuentes, pero obvian la incorporación del resto de parámetros. Esto constataría el progreso necesario del periodismo de datos para alcanzar el ideal de transparencia marcado en esta investigación. Las informaciones basadas en datos deberían ofrecer la transparencia que los propios periodistas reclaman a los gobiernos e instituciones.

Se puede concluir, por tanto, que en el desarrollo del periodismo de datos la transparencia posee un papel fundamental y que parte de su reconocimiento y credibilidad está ligada a ella. No obstante, lo que resulta evidente con esta investigación es que su implantación es aún incipiente y no se aprovechan las oportunidades que ofrecen las webs y las herramientas gratuitas en las distintas fases del proceso de producción y comunicación.

El tipo de estudio planteado impide una aproximación al tema abordado desde una perspectiva cualitativa. En futuras investigaciones se pretende completar este trabajo con entrevistas en profundidad a periodistas de datos para conocer los motivos de la opacidad de las piezas. Aún así, los hallazgos aquí plasmados sirven para reflejar cómo ha sido el trabajo desarrollado en los últimos años, qué características ostentan y cuáles son las tendencias futuras relacionadas con el ejercicio del periodismo de datos. Estas publicaciones premiadas marcan las tendencias futuras en esta práctica por innovar e integrar variedad de recursos. Asimismo, a diferencia de los estudios anteriores, aquí se plantea una metodología robusta para

## ARTÍCULOS DE INVESTIGACIÓN

el análisis de la transparencia en su conjunto, permitiendo evaluar su estado y observar su desarrollo, y se establecen los requisitos para alcanzar un grado óptimo de transparencia.

## Referencias

- Alexander, S. & Vetere, C. (2011). Telling the data story the right way. *Healthcare financial Management*, 65(10), 104-110. Recuperado de <https://bit.ly/33dFaVe>
- Allen, S. D. (2008). The Trouble with Transparency. The challenge of doing journalism ethics in a surveillance society. *Journalism Studies*, 9(3), 323-40. <https://doi.org/10.1080/14616700801997224>
- Anderson, B. & Borges-Rey, E. (2019). Encoding the UX: User Interface as a Site of Encounter between Data Journalists and Their Constructed Audiences. *Digital Journalism*, 7(9), 1253-1269. <https://doi.org/10.1080/21670811.2019.1607520>
- Anderson, C. W. (2018). *Apostles of certainty: Data journalism and the politics of doubt*. Oxford: Oxford University Press.
- Appelgren, E. & Salaverría, R. (2018). The promise of the transparency culture. A comparative study of Access to public data in Spanish and Swedish newsrooms. *Journalism Practice*, 12(8), 986-996. <https://doi.org/10.1080/17512786.2018.1511823>
- Beiler, M., Irmer, F. & Breda, A. (2020). Data Journalism at German Newspapers and Public Broadcasters: A Quantitative Survey of Structures, Contents and Perceptions. *Journalism Studies*, 21(11), 1571-1589. <https://doi.org/10.1080/1461670X.2020.1772855>
- Campos Domínguez, E. & Redondo García, M. (2015). Meta periodismo y transparencia informativa en el periodismo del siglo XXI. *OBETS, Revista de Ciencias Sociales*, 10(1), 185-209. <https://doi.org/10.14198/OBETS2015.10.1.07>
- Coddington, M. (2015). Clarifying Journalism's Quantitative Turn. *Digital Journalism*, 3(3), 331-348. <https://doi.org/10.1080/21670811.2014.976400>
- Cortés del Álamo, H., Luengo Cruz, M. & Elías, C. (2018). Periodismo de datos y transparencia al margen de los grandes medios, un estudio comparativo de Civio y Propública. *Revista ICONO14. Revista científica de Comunicación y Tecnologías emergentes*, 16(2), 66-87. <https://doi.org/10.7195/ri14.v16i2.1177>

- Deuze, M. (2003). The Web and Its Journalism: considering the consequences of different type of news media online. *New Media & Society*, 5(2), 203-30. <https://doi.org/10.1177/1461444803005002004>
- Díez Garrido, M. & Campos Domínguez, E. M. (2018). Los periodistas ante la transparencia en España: valoración y uso de la apertura informativa. *Revista Española de la Transparencia*, 7, 49-69. Recuperado de <https://bit.ly/3h0ioI7>
- Felle, T. (2016). Digital watchdogs? Data reporting and the news media's traditional 'fourth estate' function. *Journalism*, 17(1), 85-96. <https://doi.org/10.1177/1464884915593246>
- Fink, K. & Anderson, C. (2015). Data Journalism in the United States. *Journalism Studies*, 16(4), 467-481. <https://doi.org/10.1080/1461670X.2014.939852>
- García, A. M. (2014). Open government, open data, big data y transparencia: la información como nexo de unión. *COMeIN*, 39. Recuperado de <https://bit.ly/3kUEzOo>
- Gehrke, M. (2020). Transparency as a key element of data journalism. Perceptions of Brazilian professionals". En: *Computation + Journalism Symposium*. Boston: USA. Recuperado de <https://bit.ly/3l3jhy5>
- Hammond, P. (2017). From computer-computer-assisted to data-driven: Journalism and Big Data. *Journalism*, 18(4), 408-424. <https://doi.org/10.1177/1464884915620205>
- Hayes, A., Singer, J. B & Ceppos, J. (2007). Shifting Roles, Enduring Values: the credible journalist in a digital age. *Journal of Mass Media Ethics*, 22(4), 79-262. <https://doi.org/10.1080/08900520701583545>
- Horky, T. & Pelka, P. (2017). Data Visualization in Sports Journalism. *Digital Journalism*, 5(5), 587-606. <https://doi.org/10.1080/21670811.2016.1254053>
- Howard, A. B. (2014). *The art and science of data-driven journalism*. New York: Tow Center for Digital Journalism. <https://doi.org/10.7916/D8Q531V1>
- Jamil, S. (2019). Increasing Accountability Using Data Journalism: Challenges for the Pakistani Journalists. *Journalism Practice*, 1-22. <https://doi.org/10.1080/17512786.2019.1697956>
- Karlsson, M. (2010). Rituals of Transparency. *Journalism Studies*, 11(4), 535-545. <https://doi.org/10.1080/14616701003638400>

## ARTÍCULOS DE INVESTIGACIÓN

- Karlsson, M. (2020). Dispersing the opacity of transparency in Journalism on the appeal of different forms of transparency to the general public. *Journalism Studies*, 21(13), 1795-1814. <https://doi.org/10.1080/1461670X.2020.1790028>
- Karlsson, M. & Clerwall, C. (2018). Transparency to the Rescue? Evaluating citizens' views on transparency tools in journalism. *Journalism Studies*, 19(13), 1923-1933. <https://doi.org/10.1080/1461670X.2018.1492882>
- Karlsson, M., Clerwall, C. & Nord, L. (2017). Do not stand corrected: transparency and users' attitudes to inaccurate news and corrections in online journalism. *Journalism & Mass Communication Quarterly*, 94(1), 148-167. <https://doi.org/10.1177/1077699016654680>
- Kennedy, H., Weber, W. & Engebretsen, M. (2020). Data visualization and transparency in the news. En: Engebretsen, M. & Kennedy, H. (eds.). *Data Visualization in Society*. Amsterdam: Amsterdam University Press.
- Kovach, B. & Rosentiel, T. (2001). *The Elements of Journalism. What news people should know and the public should expect*. New York: Crown Publishers.
- La-Rosa, L. & 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(11), 1208-1229. <https://doi.org/10.4185/rlcs-2016-1142>
- Lizcano, J. (2012). Transparencia. *Eunomía. Revista en Cultura de la Legalidad*, 3, 160-166. Recuperado de <https://bit.ly/3pTVFQ6>
- Loosen, W., Reimer, J. & Schmidt, F. (2017). Data-driven reporting: An on-going revolution? An analysis of projects nominated for the Data Journalism Awards 2013-2016. *Journalism*, 00(0), 1-18. <https://doi.org/10.1177/1464884917735691>
- Lowrey, W. & Hou, J. (2018). All forest, no trees? Data journalism and the construction of abstract categories. *Journalism*, 1-17. <https://doi.org/10.1177/1464884918767577>
- Manfredi, J. L. (2014). Buenas y malas noticias sobre la ley de transparencia". *Cuadernos de periodistas*, 27, 72-80. Recuperado de <https://bit.ly/3fGnuH7>
- Martinisi, A. (2013). "Data Journalism and its role in Open Government". En: *International Scientific Conference e-Governance, Research and Educational Centre for e-Governance*. Universidad Técnica de Sofía, Sofía (Bulgaria), junio de 2013. Recuperado de <https://bit.ly/33dCw21> Sofía, Sofía (Bulgaria), junio de 2013. Recuperado de <https://bit.ly/33dCw21>

- Meyer, P. (1991). *The New Precision Journalism*. Indianapolis: Indiana University Press.
- Paraise, S. & Darigal, E. (2012). Data-driven journalism and public good: “Computer-assisted-reporters” and “programmer-journalists” in Chicago. *New Media & Society*, 15(6), 853-871. <https://doi.org/10.1177/1461444812463345>
- Philips, A. (2010). Transparency and the New Ethics of Journalism. *Journalism Practice*, 4(3), 373-382. <https://doi.org/10.1080/17512781003642972>
- Plaisance, P. L. (2000). The Concept of media Accountability Reconsidered. *Journal of Mass Media Ethics*, 15(4), 257-268. [https://doi.org/10.1207/S15327728JMME1504\\_5](https://doi.org/10.1207/S15327728JMME1504_5)
- Porlezza, C. & Splendore, S. (2019). From Open Journalism to Closed Data: Data Journalism in Italy. *Digital Journalism*, 7(9), 1230-1252. <https://doi.org/10.1080/21670811.2019.1657778>
- Porter, T.M. (1996). *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton: Princeton University Press.
- Singer, J. (2010). Norms and the network: Journalist ethics in a shared media space. En: Meyers, C. (Ed.). *Journalism Ethics: A philosophical approach*. (pp. 227-239). Oxford: Oxford University Press.
- Solop, F. I. & Wonders, N. A. (2016). Data Journalism Versus Traditional Journalism in Election Reporting: An Analysis of Competing Narratives in the 2012 Presidential Election. *Electronic News*, 10(4), 203-223. <https://doi.org/10.1177/1931243116656717>
- Stalph, F. (2017). Classifying Data Journalism. A content analysis of daily data-driven stories. *Journalism Practice*, 12(1), 1332-1350. <https://doi.org/10.1080/17512786.2017.1386583>
- Tal, A. & Wansink, B. (2016). Blinded with science: Trivial graphs and formulas increase ad persuasiveness and belief in product efficacy. *Public Understanding of Science*, 25(1), 117-125. <https://doi.org/10.1177/0963662514549688>
- Tandoc, JR. & Oh, S. (2017). Small Departures, Big Continuities? Norms, values and routines in The Guardian’s big data journalism. *Journalism Studies*, 18(8), 997-1015. <https://doi.org/10.1080/1461670x.2015.1104260>
- Vos, T. P. & Craft, S. (2017). The Discursive Construction of Journalistic Transparency. *Journalism Studies*, 18(12), 1505-1522. <https://doi.org/10.1080/1461670x.2015.1135754>

## ARTÍCULOS DE INVESTIGACIÓN

- Weinberger, D. (2009). Transparency: The new objectivity. *Knowledge Management World*. Recuperado de <https://bit.ly/3fw1Vc2>
- Young, M., Hermida, A. & Fulda, J. (2018). What Makes for Great Journalism? A content analysis of data journalism awards finalists 2012-2015. *Journalism Practice*, 12(1), 115-135. <https://doi.org/10.1080/17512786.2016.1270171>
- Zamith, R. (2019). Transparency, Interactivity, Diversity and Information Provenance in Everyday Data Journalism. *Digital Journalism*, 7(4), 470-489. <https://doi.org/10.1080/21670811.2018.1554409>
- Zhang, S. & Feng, J. (2019). A step forward? Exploring the diffusion of data journalism as journalistic innovations in China. *Journalism Studies*, 20(9), 1281-1300. <https://doi.org/10.1080/1461670X.2018.1513814>

## Notas

- [1] La correspondencia de los niveles de transparencia de los datos es la siguiente: 0 = No menciona la fuente ni aporta información del *dataset*; 1 = Menciona la fuente, pero no aporta información del *dataset*; 2 = No menciona la fuente, pero aporta información del *dataset*; 3 = Menciona la fuente e información del *dataset*.
- [2] La correspondencia del apartado sobre metodología es el siguiente: 0 = No incluye metodología; 1 = Sí, dentro del propio texto; 2 = Sí, en un apartado específico.
- [3] El gráfico plasma las categorías de cada variable (representadas con un triángulo y en formato <variable\_categoría>) y el número asignado a cada uno de los proyectos (representados con un punto). Las variables y sus correspondientes categorías aparecen detalladas en los apartados “Material y métodos” y “Resultados”.



Esta obra está bajo una licencia de [Creative Commons Reconocimiento 4.0 Internacional](https://creativecommons.org/licenses/by/4.0/).