

Universidad Miguel Hernández de Elche
Facultad de Ciencias Sociales y Jurídicas de
Elche

Titulación de Periodismo

Trabajo Fin de Grado

Curso Académico 2021-2022



***Los lenguajes de programación y las herramientas
digitales en el periodismo de datos en España***

***Programming languages and digital tools in data
journalism in Spain***

Alumno: Enrique Ribera Carbonell

Tutor: Félix Arias Robles

Resumen

El periodismo de datos es actualmente una de las especialidades con mayor proyección en la industria. Los medios cada vez recurren más a esta especialidad para contar historias de una manera más objetiva y esto ha aumentado con la pandemia provocada por la COVID-19 debido a la gran cantidad de datos que se extraen (contagiados, vacunados, fallecidos...). Este trabajo busca descubrir los lenguajes de programación que utilizan y los casos en los que prefieren utilizar herramientas alternativas. También investiga sobre el perfil profesional y la calidad de la formación que ofrecen las universidades en España en esta modalidad. Para esto hemos realizado una entrevista a un miembro de cada medio de comunicación nacional que trabaja con los datos (n = 10). Los resultados muestran que R y Python son los más utilizados para la obtención, limpieza y análisis de los datos, mientras que JavaScript es el preferido para la visualización. En cuanto a las herramientas, las más valoradas son Excel, Google SpreadSheets y OpenRefine para la obtención, limpieza y análisis; para la visualización, predomina el uso de Datawrapper y Flourish. Finalmente se observa que las características que debe poseer un periodista que trabaja con datos son la capacidad de entrevistar a los números, entender las bases de datos y no tenerle miedo a la tecnología. La investigación refleja que en los grados de las universidades españolas no se ofrece la formación necesaria, pero que en los másteres sí que se enseñan estas habilidades. En ambos casos se tiende a ser optimistas de cara al futuro ya que cada vez se van ofreciendo más conocimientos en los grados y existe mayor oferta de másteres de esta especialización.

Palabras clave: Periodismo de datos, lenguajes de programación, análisis, visualización, herramientas

Abstract: Data journalism is considered one of the most important specialties in today's journalism. Media outlets draw on this specialty to tell stories in a more objective and profound way. Its prominence has increased with the pandemic situation caused by COVID-19 because of the great amount of data that can be extracted (vaccinated, deceased, infected...). This study aims to explore which programming languages are frequently used and in what cases data journalists prefer employing alternative tools. This

study also tries to research the professional profile required to develop this sort of contents and evaluate the quality of the professional training offered by Spanish universities on this modality. For this, we have conducted an interview with a member of every national media organizations that post data stories online (n=10). The results show that R and Python are the most used programming languages to collect the data, while JavaScript is preferred to elaborate the visualizations. In relation to the alternative tools, the most valued ones are Excel, Google Spreadsheets and OpenRefine to obtain, clean and analyze data, while DataWrapper and Flourish predominate for the visualizations. Finally, we observe that the characteristics that every data journalist must possess are the ability to "interview" the numbers, understand databases and not being afraid of technology. The research reflects that Bachelor degrees don't offer enough training, while Master's degrees do. In both cases, the situation seems to be improving, as knowledge offered in Bachelor's degree is growing, and there is a greater offer of master's degrees on this specialization.

Keywords: Data journalism, programming languages, analysis, visualization, tools



Índice

1. Introducción.....	4
1.1 Objetivos del estudio:.....	5
1.2 Hipótesis:.....	5
2. Estado de la cuestión	7
2.1 El concepto de periodismo de datos	7
2.2 La tecnología en el periodismo de datos	8
2.3 Procesos en el periodismo de datos.....	9
2.4 Perfil profesional del periodista de datos	11
3. Metodología.....	12
3.1 Selección:	12
3.2 Trabajo de campo	14
4. Resultados.....	16
4.1 Preferencias de los periodistas en los lenguajes de programación.....	16
4.2 Usos de lenguajes de programación	17
4.3 Usos de los lenguajes de programación	18
4.4 Usos de herramientas alternativas	21
4.5 Perfil profesional según los periodistas.....	23
4.6 Visión general de los profesionales sobre la programación.....	25
4.7 Equipos de datos.....	27
5. Conclusiones.....	29
5.1 Programación en periodismo.....	29
5.2 Herramientas alternativas a la programación más utilizadas	31
5.3 Perfil profesional	32
6. Anexos	35
6.1 Periodistas de datos de los medios de comunicación seleccionados.....	35
6.2 Librerías	37
6.2.1 Librerías JavaScript	37
6.2.2 Librerías R	37
6.2.3 Librerías Python	37
6.2.4 Librerías Ruby	38
7. Bibliografía.....	39

1. Introducción

Esta investigación tiene como objeto de estudio analizar las diferentes aplicaciones de los lenguajes de programación para la producción de contenidos basados en el periodismo de datos. Pretende explorar el panorama nacional (España) y descubrir las diferentes formas de trabajar con las nuevas tecnologías para desarrollar contenidos a partir de grandes cantidades de datos, que no serían posible diseccionarlos sin la ayuda de los lenguajes de programación, o con herramientas.

El periodismo de datos no es una disciplina nueva, pero desde 2009 ha ido aumentando hasta convertirse en uno de los pilares de los medios de comunicación (Polaino y Vivar, 2020). Este incremento se debe a las herramientas que permiten tratar datos masivos (Big Data) y a las características de la Web 2.0 con las que internet se transforma a un modelo social y participativo, dando pie al conocido como “periodismo digital”, que aprovecha las funciones del hipertexto, multimedialidad y actualización constante (Barceló, 2020).

Estos diferentes tipos de periodismo se enmarcan en el denominado “periodismo computacional”, que hace referencia a los sistemas que automatizan parte de los procesos de creación y de personalización de piezas periodísticas. En cuanto al periodismo de datos, permite limpiar, procesar y visualizar de forma atractiva grandes cantidades de datos, además de poder obtener datos directamente de páginas web aplicando fórmulas de raspado (técnica para extraer la información de las bases de datos de las páginas web a través de la estructura HTML de la página) (Vállez y Codina, 2018).

Estos avances tecnológicos han obligado a los periodistas a adaptarse y a formarse para poder dar valor añadido a sus historias aprovechando las ventajas que ofrecen las herramientas mencionadas, por ello veremos qué perfiles profesionales solicitan los medios para este tipo de especialidad.

Para finalizar, analizaremos los diferentes productos, elaborados a partir de estas tecnologías, que han lanzado los principales medios de comunicación en el último año (2020). Hay que tener en cuenta que este año ha habido un incremento en el uso de estadísticas por parte de los medios debido a las grandes cantidades de cifras producidas

por la pandemia provocada por la COVID-19 (infectados, fallecidos, test realizados, vacunas inoculadas...).

1.1 Objetivos del estudio:

Los principales objetivos de este estudio son comprobar los lenguajes de programación que utilizan los medios para realizar piezas más innovadoras en el periodismo de datos y las herramientas alternativas que ayudan a realizar tareas complejas sin tener que dominar la programación. Conocer estos lenguajes y/o herramientas no es común en los periodistas, por eso también se investiga el perfil profesional de los que se dedican a contar historias a través de datos. Los objetivos concretos de esta investigación, por lo tanto, serían los siguientes:

- O 1. Elaborar una lista con los lenguajes de programación más utilizados.
- O 2. Descubrir los usos que les otorgan los medios de comunicación generalistas en España a los lenguajes de programación.
- O 3. Descubrir qué librerías utilizan más dependiendo del lenguaje utilizado.
- O 4. Conocer las herramientas alternativas a los lenguajes de programación más utilizadas.
- O 5. Detectar los perfiles profesionales demandados para esta modalidad del periodismo.

1.2 Hipótesis:

Según los conocimientos previos a la realización de este estudio, los lenguajes de programación más empleados en las tareas de periodismo de datos son R y Python para el proceso de análisis de datos y JavaScript para la visualización. Por otra parte las herramientas que más se utilizan son Excel, DataWrapper, Flourish y Tableau. El periodista de datos debe conocer las principales características de las bases de datos y tener conocimientos primarios en estadística. De ahí que las hipótesis de esta investigación sean las siguientes:

- H 1. Pese a que muchos medios utilizan herramientas prediseñadas, el lenguaje de programación más utilizado es R, seguido de Javascript y Python, en ese orden.
- H 2. Los medios de comunicación utilizan más estos lenguajes para obtener grandes bases de datos, limpiarlas y visualizarlas.
- H 3. En R la librería más utilizada es el conjunto “tidyverse”. En JavaScript “D3” y en Python “Pandas”.
- H 4. Las alternativas más utilizadas son Excel, Flourish, DataWrapper y Tableau.
- H 5. El perfil demandado es el periodista con formación en técnicas de obtención de bases de datos y su posterior limpieza y visualización. Estos periodistas deberán tener conocimientos sobre estadística para manejar los datos de forma óptima y manejo de herramientas tecnológicas.



2. Estado de la cuestión

2.1 El concepto de periodismo de datos

El periodismo de datos se nutre del periodismo computacional, entendido como el uso de los sistemas que automatizan parte de los procesos de creación y de personalización de contenidos periodísticos. Se forma por dos vías: la continuista, que se trata de la utilización de la tecnología para agilizar y optimizar las tareas realizadas por los periodistas; y la disruptiva, que implica nuevas formas de hacer periodismo, cuando las redacciones incorporan el uso intensivo de la tecnología en fases de sus procesos que hasta hace poco eran impensables, como la inteligencia artificial aplicada a la creación automática de noticias (Vállez y Codina, 2018).

El periodismo de datos se incluye dentro de la vía continuista ya que la utilización de datos para la producción de contenidos periodísticos se lleva a cabo desde siempre (Crucianelli, 2013), pero no fue hasta la década de los 60 en la era necesaria la ayuda de una computadora para la gestión de estos datos en algunos medios de los EEUU, pero el “periodismo basado en datos” contemporáneo mejora la forma en que el periodismo puede contribuir a la democracia, especialmente en un momento en que los gobiernos publican un número creciente de conjuntos de datos (Parasie y Dagiral, 2013). En España, con la entrada en vigor en 2014 de la Ley de Transparencia del Gobierno de España y el emergente concepto del Big data, muchos periodistas se vienen especializando en el Periodismo de Datos (Vivar, 2017).

Dada la evolución de la disciplina, que aún se encuentra en desarrollo, existen diferentes definiciones de varios expertos: el periodismo de datos es una forma de periodismo de investigación que utiliza la informática para buscar, contrastar, visualizar, combinar datos y proporcionar información de diferentes fuentes, públicas o privadas (Flores y Cebrián, 2012); qué es la especialidad o práctica del periodismo que utiliza como materia prima y fundamental datos (Peiró y Guallar, 2013); el periodismo de datos se diferencia del resto por “las nuevas posibilidades que se abran cuando se combina el tradicional olfato periodístico y la capacidad de contar una historia convincente, con la magnitud y el alcance de la información digital disponible (Bradshaw, 2012).

2.2 La tecnología en el periodismo de datos

A medida que la tecnología evoluciona ha sido posible rastrear y analizar bases de datos con el fin de contextualizar hechos informativos y comienza a hablarse de database journalism, Computer Assisted Journalism (CAJ) o de CAR (Computer Assisted Reporting), modalidad del periodismo de precisión en la que se utilizan los ordenadores para examinar las bases de datos y descubrir asociaciones o correlaciones estadísticas en todo tipo de documentos informatizados (Ferrerías, 2016). El periodismo de precisión es una modalidad formulada por el profesor Philip Meyer, involucra el uso de métodos de la investigación social aplicados al periodismo tales como la investigación cualitativa (análisis de casos, grupos focales, análisis del discurso, etc.) y la cuantitativa. (Crucianelli, 2013). Su despegue definitivo como disciplina vendría de la mano del profesor Philip Meyer en 1973 con la publicación de la primera edición de su libro Precision Journalism (Ferrerías, 2016).

Los periodismos de precisión y de investigación comparten muchos puntos en común, pero se trata de dos ámbitos independientes que en muchas ocasiones comparten sus métodos y objetivos para obtener resultados de interés común. El periodismo de investigación siempre ha respondido al objetivo de desvelar unos hechos noticiosos cuyos principales protagonistas no están dispuestos a que sean conocidos públicamente. Aquí es donde coincide con el periodismo de precisión. Cuando se dan tales circunstancias el periodismo de precisión se convierte en un periodismo de investigación mediante nuevos métodos, habitualmente no conocidos o no utilizados por los profesionales de la investigación convencional procedentes de métodos de las ciencias sociales y singularmente del campo estadístico y el análisis informático (Dader, 2002).

Aunque puede que no haya una diferencia entre los objetivos y técnicas del periodismo de precisión y el de datos, la aparición de la etiqueta de 'periodismo de datos' a principios del siglo indica una nueva fase en la que el volumen de datos que está disponible gratuitamente en línea, en combinación con sofisticadas herramientas centradas en el usuario, la autopublicación y herramientas *crowdsourcing*, permite a más gente a trabajar con más datos con facilidad (Bounegru, 2012).

El Big Data nace del interés por las grandes cantidades de datos disponibles en internet gracias al uso masivo de los denominados medios sociales (Facebook, Twitter, blogs, etc.) y de otras fuentes textuales de información (medios de comunicación online, webs oficiales, etc). Se ha intentado vincular el big data solo con el tamaño de los datos (por ejemplo, en los papeles de Panamá se usaron 2,6 terabytes), esta dimensión no es suficiente para caracterizarlo, ya que el concepto se refiere fundamentalmente a volúmenes masivos y complejos de información tanto estructurada como no estructurada, que es recogida durante cierto período de tiempo y que requiere de métodos computacionales para extraer conocimiento (Arcila, Barbosa y Cabezuelo, 2016). Otros conceptos importantes ligados al estudio del big data también aluden a su intencionalidad y utilidad (Murphy y Barton; 2014).

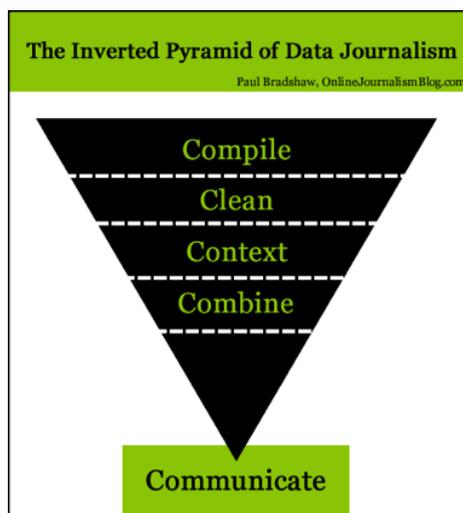
Existen tres retos asociados al fenómeno big data: problemas tecnológicos asociados al almacenamiento, seguridad y análisis de los siempre crecientes volúmenes de datos; valor comercial que puede ser añadido a través de la generación de perspectivas más efectivas; impactos sociales, particularmente las implicaciones para la privacidad personal (Nunan y Di Domenico, 2013).

Estos retos están asociados a tres cambios en el paradigma: mayor importancia de la disponibilidad y acceso de los datos; aceptación de niveles de imprecisión y desorden en los datos; centrarse más en las correlaciones, en vez de buscar constantemente la causalidad (Mayer y Cukier, 2013).

2.3 Procesos en el periodismo de datos

Por tanto, se puede calificar el periodismo de datos como una nueva modalidad o disciplina informática en la que se dan tres elementos fundamentales: acceso a los datos, su tratamiento y su publicación; cuyo denominador común son las nuevas herramientas y tecnologías. Elementos que conforman, a su vez, las fases de trabajo en esta disciplina. Precisamente Bradshaw publicó un diagrama, con forma de pirámide invertida, para representar las fases del proceso de producción. La forma de pirámide invertida muestra la cantidad de datos con la que se trabaja en los primeros procesos y se va enfocando cada vez que se va profundizando en las fases (Bradshaw, 2011).

Imagen1: Pirámide invertida de Periodismo de Datos.



Fuente: <http://onlinejournalismblog.com>

La compilación de datos, o la obtención de datos, se refiere a las fuentes por las que se consiguen y existen dos tipos de fuentes: públicas y privadas. Se puede acceder a ellas mediante peticiones de información, búsquedas avanzadas o con técnicas de scraping (raspado). También es posible obtener datos tomando información de APIs o a través de encuestas, formularios online o *crowdsourcing*.

La limpieza hace referencia a comprobar la calidad de estos y unificar sus formatos. Se puede realizar mediante lenguajes de programación estadísticos como R o con otros lenguajes como Python o Ruby. También se puede realizar mediante herramientas como OpenRefine, Excel o Google SpreadSheets.

El análisis de los datos (“context”) está asociada con la interpretación de estos para poder entender mejor su significado y va muy ligado a la combinación de los datos con otros o con datos geoespaciales y mapas.

Por último, tenemos la comunicación o la visualización de estos, es el momento en el que se exponen los resultados de una forma visual como con tablas, gráficas, mapas o con otros formatos como textos o aplicaciones. (Bradshaw, 2011). Actualmente existen muchas herramientas que permiten una sencilla creación de estas visualizaciones como DataWrapper o Flourish y librerías muy avanzadas como D3 de JavaScript.

Una recomendación que realizan muchos periodistas es la conveniencia de publicar además los datos brutos y explicar las metodologías utilizadas en el proceso; con el fin de fomentar la transparencia, pero también la participación de los públicos (Ferrerías, 2016).

2.4 Perfil profesional del periodista de datos

La delimitación de funciones y perfiles de los profesionales de la información vuelven a ser revisados, como en anteriores ocasiones, ante la evolución de la actividad periodística a lo largo de su historia (Sánchez, Campos, Berrocal, 2015). En esta especialidad se necesita un profesional que puede estar formado en documentación, periodismo o cualquier otra disciplina, que posee un aditivo de conocimiento necesario para la gestión de datos y clasificación de información, así como para la búsqueda, procesamiento, interpretaciones concluyentes, análisis de datos y capacidad para crear y publicar información con una narrativa propia que se desprende del propio género utilizado. Deberá manejar grandes cantidades de información, deberá clasificar y después, introducir esos datos en las aplicaciones tecnológicas de uso, por donde considere que su información tendrá más impacto. Además, deben tener conocimientos sobre Estadística, Programación, Diseño, Sociología, Psicología, además del Periodismo (Flores y Cebrián, 2012).

Para el Periodismo de Datos hace falta entrenamiento y aprendizaje constantes y conocimientos de materias generalmente alejadas de las preferencias del periodista convencional, se debe conocer el funcionamiento de la administración pública, dominio del inglés para entender las herramientas más utilizadas, ya que no tienen su versión en español, habilidades en búsquedas en la web más avanzadas de las habituales, conocer los principios básicos de programas de análisis de datos (Excel, Google SpreadSheets...), conocimientos de matemática básica y de estadística para poder entrevistar a los números, así como de visualización (Crucianelli, 2012). Además, añade que en esta tarea solo brillan los perseverantes.

Otros profesionales entienden que los conocimientos en programación son fundamentales, así como encontrar historias dentro de los datos (Ferrerías, 2013).

3. Metodología

Esta investigación se basa en la realización de entrevistas a periodistas de datos de medios de comunicación generalistas que publican en internet (sean nativos digitales o tengan matriz analógica) y que publiquen utilizando lenguajes de programación o herramientas alternativas. Estas entrevistas sirven para que los entrevistados muestren qué lenguajes y qué librerías utilizan y, si no utilizan lenguajes de programación, las herramientas alternativas que utilizan para la elaboración de piezas basadas en datos. Las entrevistas se han realizado mediante el formato de entrevista semiestructurada: tienen el mismo patrón de preguntas, pero son flexibles ya que cada periodista trabaja de una manera determinada.

3.1 Selección:

La selección de los medios se basa en la audiencia que han tenido en los dos primeros meses de 2021 sus medios digitales y la cantidad de seguidores con los que cuentan en Redes Sociales.

De estos nos hemos quedado solo con los que tienen un equipo dedicado al periodismo de datos y se han valorado la cantidad de periodistas de datos con los que cuentan en su plantilla.

Medio	Matriz	Twitter	Facebook	Instagram	Audiencia	Periodistas	PUNTOS
Maldita	Digital	117500	63000	31300	1.266.222	8	10,1
El País	Prensa	8000000	5300000	1100000	19.095.008	5	9,8
Civio	Digital	44900	10300		83.710	6	7,5
El Confidencial	Digital	910900	1000000	142000	19.263.246	5	7,2
RTVE	Audiovisual	1200000	778000	263000	9.722.684	5	7,1
Newtral	Digital	187700	33500	40300	401.091	5	6,3
Storydata	Servicio comunicación	2173	156	734		4	5,0
El Mundo	Prensa	3900000	2600000	504000	22.732.110	2	4,6
eldiario.es	Digital	1200000	576000	153000	10.691.610	3	4,5
La sexta	Audiovisual	1000000	479000	123000	9.632.075	2	3,1
Datadista	Digital	26200	8200	855		2	2,5
La Vanguardia	Prensa	1200000	4300000	534000	34.972.749	0	2,4
Telecinco	Audiovisual	1700000	1600000	1000000	10.899.626	0	1,7
ABC	Prensa	525500	1600000	363000	20.059.749	0	1,3
Antena 3	Audiovisual	1400000	1900000	402000	10.130.797	0	1,2
20 Minutos	Prensa	1400000	2900000	197000	14.858	0	0,9
El Español	Digital	427700	733000	90700	20.726.488	0	0,9
El Periódico	Prensa	696000	823000	95400	12.739.528	0	0,7
OkDiario	Digital	331000	868000	164000	10.729.767	0	0,7
Cuatro	Audiovisual	1100000	839000	236000	4.214.905	0	0,6
La Razón	Prensa	525500	446000	117000	8.426.492	0	0,5

Fuente: Elaboración propia

Después de esta criba se quedaron 11 medios (Maldita, El País, Civio, El Confidencial, RTVE, Newtral, Storydata, El Mundo, elDiario.es, La Sexta y Datadista) y se escogió a un periodista que trabajase en el medio. El objetivo inicial fue contactar con el responsable de la sección de datos de cada medio. En caso de no ser posible entrevistarlos, se ha pasado a otro periodista que integre el equipo de datos del medio.

De los medios seleccionados, se han podido realizar 10 entrevistas ya que no se ha podido concretar una entrevista con un medio (El Mundo). Para contactar con los entrevistados se ha recurrido al Gmail (personal o corporativo) si esta publicado. Con los que no tienen el Gmail público se ha contactado con ellos a través de redes sociales como Twitter o LinkedIn.

3.2 Trabajo de campo

Las preguntas realizadas durante la entrevista están divididas en tres bloques con un total de 21 preguntas más tres preguntas con un formato de escala de Likert. En el primero se busca conocer las opiniones sobre los lenguajes de programación en general y en las situaciones en las que recurren a ellos. Este bloque cuenta con tres preguntas sobre su opinión general sobre la programación en el periodismo y los motivos por los que recurren a ellos.

El segundo bloque se centra en descubrir qué lenguajes de programación utilizan para cada parte del proceso (obtención de datos, limpieza, análisis y visualización) y las librerías que utilizan, así como las herramientas elegidas como alternativa a estos lenguajes. Se ha conformado con un total de 10 preguntas. Cuatro para descubrir los lenguajes de programación que utilizan en cada parte del proceso, otras cuatro para saber las herramientas alternativas en estas partes, una para descubrir las librerías utilizadas y otra para saber la percepción del lenguaje que es más utilizado durante su trabajo.

El último bloque cuenta con 8 preguntas y se centra en conocer la percepción que tienen sobre el perfil profesional necesario para ser periodista de datos, los equipos de esta especialidad de sus medios y sobre la formación de estos conocimientos tanto en grados y másteres como en sus propios medios.

Para finalizar, se ha realizado una breve lista de tres preguntas de tipo Escala de Likert en las que los entrevistados han puesto una nota (de 0 a 10) a cada una de las afirmaciones propuestas. En un principio, esta lista de preguntas contenía un total de seis cuestiones, pero tres han sido eliminadas por centrarse en más ámbitos que el periodismo de datos y los entrevistados no han tenido conocimientos suficientes para poder contestarlas.

Las entrevistas han tenido una duración media de 35 minutos. La mayoría de las entrevistas se han realizado a través de la plataforma de videollamada *Meet* para poder compartir pantalla para la explicación de las respuestas. En caso de no poder realizarla por *Meet* se han llevado a cabo vía telefónica con la misma duración media. Para comprobar la idoneidad de las preguntas y la duración de las entrevistas se realizó un

pretest con uno de los seleccionados (Datadista). En este pretest se comprobó la necesidad de aclarar que las preguntas buscan conocer los pensamientos de los profesionales y la forma en la que trabajan.



4. Resultados

Los resultados obtenidos a través de las entrevistas realizadas nos muestran cómo se trabaja actualmente el periodismo de datos en los medios españoles. Python es el lenguaje de programación preferido, pero no es el más utilizado. Las herramientas con las que trabajan alternativas al lenguaje de programación más utilizadas son Datawrapper, Flourish, Excel o Google Spreadsheets y OpenRefine. También su visión sobre el perfil necesario para trabajar con datos.

4.1 Preferencias de los periodistas en los lenguajes de programación

Tabla 1: Mejores lenguajes de programación para crear contenido periodístico

Lenguaje de programación	Nº	%
Python	8	80
JS	7	70
R	6	60
Ruby	1	10
PHP	1	10
SQL	1	10

Fuente: Elaboración propia

Podemos observar que los profesionales valoran a Python como el lenguaje de programación óptimo para crear contenidos periodísticos (80%), muy seguido por JavaScript (70%) y por R (60%). Más alejados se encuentran Ruby, PHP o SQL que solo han recibido el 10% de las respuestas. Algunos entrevistados añaden que no hay uno mejor que otro. Mientras que otros explican que el lenguaje que más domines será el mejor para cada uno. Un encuestado explica que R es una alternativa muy buena por la amplia comunidad que tiene ya que esta actualiza constantemente y es una ayuda para cuando tienes problemas con el código.

4.2 Usos de lenguajes de programación

Tabla 2: Motivos para usar lenguajes de programación

Motivo para la utilización de la programación	Nº	%
Grandes cantidades de datos	7	70
Automatización del proceso	4	40
Extracción de información	3	30
Complejidad de la tarea	2	20
Visualizaciones especiales	2	20
En función del tiempo	1	10
Cruzar bases de datos	1	10

Fuente: Elaboración propia

Los resultados muestran que los periodistas se apoyan de los lenguajes de programación cuando se encuentran con grandes cantidades de datos (70%). Los siguientes motivos son la necesidad de automatizar los procesos (40%), necesidad de extraer la información (30%), la complejidad de las tareas (20%) y cuando se requieren visualizaciones especiales (20%). Por último, los motivos que menos se han mencionado son la disponibilidad de tiempo (10%) y cuando tienen que cruzar bases de datos (10%). Solo tres entrevistados han dado solo una respuesta en esta pregunta, cuatro han dado dos respuestas y otros tres han contestado con tres respuestas.

Tabla 3: Motivos para no usar lenguajes de programación

Motivo para no utilizar la programación	Nº	%
Poco tiempo	4	40
Poca cantidad de datos	3	30
Cuando las herramientas alternativas pueden suplir a la programación	1	10
Datos bien estructurados y sencillos	1	10

Fuente: Elaboración propia

Los periodistas valoran mucho el tiempo del que disponen para negar la utilización de los lenguajes de programación (40%), seguido por la escasa cantidad de datos necesarios (30%). Por último, mencionan las veces que puedes realizar la misma tarea con otras herramientas (10%) y cuando los datos obtenidos están bien estructurados (10%). Sin embargo, un periodista no piensa que sea en función de ninguno de los motivos anteriores, si no que para él “depende del tema”.

4.3 Usos de los lenguajes de programación

Tabla 4: Lenguajes de programación que más utilizan en el medio para el que trabajan

Lenguajes de programación	Nº	%
JS	6	60
Python	3	30
R	4	40
Ruby	1	10
Ninguno	1	10

Fuente: Elaboración propia

Los resultados muestran que el lenguaje más utilizado por los medios es JavaScript (60%), con una diferencia notable con R (40%) y con Python (30%), que pese a ser el mejor valorado para la creación de contenido no es el que más se utiliza. Por su parte, solo un medio utiliza Ruby (10%). También hay un medio (LaSexta) que no trabaja con ningún lenguaje de programación (10%).

Hay cinco medios que trabajan con un solo lenguaje (dos con JS, dos con R y uno con Ruby), tres que trabajan con dos y uno que llega a trabajar hasta con tres lenguajes (StoryData).

Hay que añadir que los lenguajes de marcado (HTML) y lenguajes de diseño (CSS) han sido muy mencionados (30%), pero al no ser lenguajes de programación no pueden entrar en este análisis.

Tabla 5: Librerías utilizadas en JavaScript

Librerías JS	Nº	%
D3	9	100
Vue.js	1	11
Leaflet	1	11
Node.js	1	11

Fuente: Elaboración propia

Como veremos en las siguientes tablas, JS es el lenguaje de programación por excelencia para la visualización de datos, todos los que utilizan la programación para este fin trabajan con él. La librería más utilizada es D3, que permite hacer visualizaciones interactivas (100%). Las siguientes librerías son Vue.js, para la construcción de interfaces de usuario

y aplicaciones, Leaflet, para crear mapas interactivos, y Node.js, que es un entorno de ejecución (11%).

Tabla 6: Librerías utilizadas en R

Librerías R	Nº	%
Tidyverse	3	100
Googlesheets	3	100
Countrycode	1	33
Rmapshaper	1	33
Data.table	1	33

Fuente Elaboración propia

Las librerías más empleadas por los que trabajan en R son Tidyverse, un conjunto de librerías diseñadas para la Ciencia de Datos, y Googlesheets, librería para conectar R con las hojas de Google (100%). Con menos uso se encuentran Countrycode, que sirve para estandarizar los nombres de los países de una base de datos, Rmapshaper, que permite editar y simplificar los objetos “geojson”, ”Spatial” y “sf”, y Data.table, que permite trabajar con grandes bases de datos de una manera más óptima (33%).

Tabla 7: Librerías utilizadas en Python

Librerías Python	Nº	%
Pandas	3	100
BeautifulSoup	1	33
Selenium	1	33
Geopandas	1	33
Geocoder	1	33
NumPy	1	33

Fuente: Elaboración propia

Los periodistas que apuestan por Python trabajan todos con la librería Pandas, una librería dedicada al manejo y análisis de datos (100%). Las otras librerías que gastan son: BeautifulSoup, para el raspado de páginas web, Selenium, para probar y registrar interacciones en páginas web, Geopandas, para trabajar con datos geoespaciales y Geocoder, librería especializada en geocodificación, es decir, conseguir las coordenadas desde una dirección y viceversa, y NumPy, que da soporte para crear vectores y matrices grandes multidimensionales, junto con una gran colección de funciones matemáticas (33%). Solo un medio ha contestado con una sola librería (Datadista) al considerar que solo utiliza Pandas de manera regular y que las otras librerías que utiliza es en función de las necesidades y son muy cambiantes.

Tabla 8: Librerías Ruby

Librerías Ruby	Nº	%
Nokogiri	1	100
Mechanize	1	100

Fuente: Elaboración propia

El único medio que trabaja con Ruby (Civio) considera que las librerías esenciales son Nokogiri, que sirve para leer, escribir, modificar y consultar documentos XML y HTML, y Mechanize, que se utiliza para automatizar la interacción con sitios web (100%).

Tabla 9: Lenguajes de programación más usados en cada uno de los procesos

Fase del proceso												
Obtención de datos			Limpieza de datos			Análisis de datos			Visualización de datos			
Lenguaje de programación	Nº	%	% total									
R	4	44	R	5	63	R	5	63	JS	9	100	90
Python	3	33	Python	2	25	Python	2	25				
Ruby	1	11	Ruby	1	13	SQL	1	13				
JS	1	11										

Fuente: Elaboración propia

Se puede ver cómo la obtención de datos es la fase del proceso que más lenguajes distintos se gastan (4) mientras que la visualización es la fase que menos lenguajes se gastan (1).

A la hora de obtener datos, los periodistas que usan lenguajes de programación apuestan por R (44%), seguido de Python (33%). Los menos recurridos para esta labor son Ruby y JavaScript (11%). Solo un medio no usa lenguajes de programación para obtener los datos.

Para limpiar los datos, los periodistas apuestan también por R (63%). Después le siguen Python (25%) y Ruby (13%). Hay dos medios que nunca gastan lenguajes de programación para esta tarea.

Para el análisis de datos se sigue apostando por R (63%). Con mucha diferencia se encuentra Python (25%) y SQL (13%). Hay dos medios que no utilizan lenguaje de programación para analizar los datos.

Todos los periodistas que utilizan lenguajes de programación para visualizar datos apuestan por JavaScript (100%). Únicamente un medio (LaSexta) no apuesta por la programación para visualizar datos.

4.4 Usos de herramientas alternativas

Tabla 10: Herramientas alternativas más usados en cada uno de los procesos

Fase del proceso															
Obtención de datos			Limpieza de datos				Análisis de datos				Visualización de datos				
Herramienta alternativa	Nº	%	% tot al	Herramienta alternativa	Nº	%	% tot al	Herramienta alternativa	Nº	%	% tot al	Herramienta alternativa	Nº	%	% tot al
Tabula	3	38	30	OpenRefine	7	70	70	Excel	8	80	80	Datawrapper	9	100	90
OpenRefine	2	25	20	Excel	6	60	60	Google Sheets*	6	60	60	Flourish	8	89	80
import.io	2	25	20	Google Sheets*	5	50	50	Datawrapper	1	10	10	Ilustrator	4	44	40
Table Capture	2	25	20	Calc	1	10	10	Flourish	1	10	10	Mapbox	4	44	40
Web Scraper	1	13	10					Tableau	1	10	10	Qgis	2	22	20
Adobe PDF	1	13	10					Calc	1	10	10	Infogram	1	11	10
Google Sheets*	1	13	10									uMap	1	11	10
Dataminer	1	13	10									RAWGraphs	1	11	10
Abyy Fine Reader	1	13	10												

Fuente: Elaboración propia

* Google Spreadsheets

Podemos observar que la obtención de datos es la fase que más herramientas distintas gastan los periodistas (9). Pese a esto, la mayoría de periodistas afirman que la gran cantidad de datos que obtienen es a través de “botones de descarga” de las páginas web. Mientras que la que menos herramientas distintas gastan es en la limpieza (4) y además tres de ellas son herramientas muy similares, pero de distintos creadores (Excel, Google Spreadsheets y Calc).

Cuando los periodistas no utilizan lenguajes de programación para obtener datos, recurren a Tabula (38%), OpenRefine, Import.io y Table Capture (25%). Por detrás de estas se encuentran Web Scraper, Adobe PDF, Google Spreadsheets, Dataminer y Abyy Fine Reader (13%). Dos medios nunca utilizan herramientas alternativas a los lenguajes de programación para la obtención de datos.

A la hora de limpiar datos, se apuesta mayormente por OpenRefine (70%), seguido de Excel (60%) y Google Spreadsheets (50%). Estas dos herramientas son utilizadas por la mayoría de los medios de forma indistinta. Excel se utiliza más para ciertas operaciones no admitidas en Google Spreadsheets y esta se utiliza para trabajos colaborativos. Otra herramienta que se utiliza es Calc (10%).

En el análisis de datos las dos herramientas por excelencia son Excel (80%) y Google Spreadsheets (60%). Después de estas están Datawrapper, Flourish (10%) para ver los patrones visuales, Tableau y Calc (10%).

A la hora de visualizar los datos Datawrapper (100%) y Flourish (89%) son las favoritas de los periodistas. Datawrapper va por delante por su mayor rapidez de carga y su visualización más limpia, según comentan varios entrevistados. Muchos de ellos utilizan Flourish cuando la visualización deseada no se puede realizar en Datawrapper. Solo un medio nunca utiliza Flourish. Después de estas herramientas, las más utilizadas son Illustrator y Mapbox (44%), Qgis (22%) y Infogram, uMap y RAWGraphs (11%). Solo un medio no utiliza nunca herramientas de visualización al tener plantillas de JS de fácil reutilización.

4.5 Perfil profesional según los periodistas

Tabla 11: Características de un periodista de datos

Características	Nº	%
Saber hacer preguntas a los datos	5	50%
No tener miedo a lo tecnológico	4	40%
Entender bases de datos	4	40%
Analista	3	30%
Organizado	2	20%
Conocer algo de programación	2	20%
Poder trabajar todo tipo de temas	2	20%
Más preocupados en el producto	1	10%
Saber de visualización	1	10%
Saber buscar enfoques sencillos con los datos	1	10%

Fuente: Elaboración propia

Los resultados muestran que los periodistas entrevistados ven en saber preguntar a los datos (50%) una de las facetas que más caracterizan a los trabajadores de esta especialidad. No tener miedo a la tecnología y entender las bases de datos (40%) y ser analista (30%) son las siguientes particularidades más mencionadas. Por último, los rasgos menos valorados han sido conocer la programación y trabajar todo tipo de temas (20%) y estar preocupado en el producto final, saber sobre visualización y saber buscar enfoques sencillos con los datos (10%).

Tabla 12: ¿Debe el periodista conseguir controlar la programación o simplemente debe tener unos conocimientos básicos para entender el trabajo de los especialistas?

Respuestas	Nº	%
Saber un poco	6	60%
Depende del entorno	2	20%
Dependiendo del proyecto	1	10%
Cuanto más sepas mejor	1	10%

Fuente: Elaboración propia

La mayoría periodistas entrevistados creen que un periodista necesita saber un poco de programación para poder entender el trabajo de los especialistas (60%). Otros periodistas opinan que depende del entorno de trabajo debes tener más conocimientos sobre

programación (20%). Un entrevistado explica que dependiendo del proyecto tienes que obtener mayor conocimiento (10%) mientras que otro explica que cuantos más conocimientos de programación se tengan mejor (10%).

Tabla 13: Conocimientos impartidos en estudios universitarios

Respuestas	Grados		Másteres	
	Nº	%	Nº	%
Si	0	0	60	60
No	6	60	0	0
NS/NC	1	10	0	0
Cada vez más	3	30	40	40

Fuente: Elaboración propia

Los periodistas creen que no se imparten los conocimientos necesarios en los grados para entender el periodismo de datos (60%). Otra gran parte piensa que cada vez se ofrecen más conocimientos (30%) mientras que un solo entrevistado no sabe el nivel de conocimientos que se imparten en los grados (10%) y nadie cree que sí que se den los conocimientos necesarios (0%).

A diferencia de los grados, muchos periodistas piensan que sí que se imparten en los másteres los conocimientos necesarios para controlar sobre periodismo de datos (60%). Otra gran cantidad de periodistas piensan que cada vez se van mejorando los currículos (40%).

Tabla 14: Formación ofrecida por los medios de comunicación

Respuestas	Nº	%
Sí	5	50
Autodidacta	4	40
Formación impartida entre compañeros	4	40

Fuente: Elaboración propia

La entrevista refleja que la mitad de los entrevistados creen que los medios sí que se esfuerzan por formarles (50%) pero muchos de estos piensan que para completar esa formación se tiene que hacer esfuerzo autodidacta (40%) o formación entre los propios compañeros (40%).

Tabla 15: Posibilidad de dominar la programación siendo periodista

Respuestas	Nº	%
Sí	7	70
De una sola especialidad	1	10
Con limitaciones	1	10
Difícilmente	1	10

Fuente: Elaboración propia

Los periodistas son optimistas a la hora de poder conseguir un gran nivel de programación (70%) mientras que son menos los que piensan que solo se puede saber de una sola especialidad (10%), que se puede conseguir con limitaciones (10%) y que es difícil de conseguir (10%).

4.6 Visión general de los profesionales sobre la programación

Tabla 16: Visión de los profesionales sobre la programación

Pregunta	Media	Mediana	Moda	Desviación típica
La programación es un recurso muy utilizado para la creación de contenidos periodísticos	5,5	5,5	5	2,3
Todo periodista debería saber programación	3,7	2,5	8	3,4
La programación potencia la creación de contenidos periodísticos	8,9	9,5	10	1,6

Fuente: Elaboración propia

La media de la puntuación obtenida por la pregunta “La programación es un recurso muy utilizado para la creación de contenidos periodísticos” es de 5,5 siendo esta también su mediana. La nota más obtenida es de 5. Pero esta pregunta, a pesar de tener a la gran mayoría en el aprobado, hay un medio que le ha dado un 10 (Datadista) y un medio que la ha puntuado con un 1 (Newtral), pese a esta diferencia la desviación típica es de 2,3.

La pregunta “Todo periodista debería saber programación” tiene una media de 3,7 y una mediana de 2,5. Pese a esto su moda es de un 8 (30%) por lo que hay una gran diferencia entre respuestas lo que lleva a obtener una desviación típica de 3´4. Dos medios le han puesto un 0 (El Confidencial y ElDiario.es) cosa que deja clara la diferencia de criterios en esta afirmación. Un periodista le ha dado un 6 defendiendo que “todos los periodistas de datos deberían saber algo de programación, pero la gran mayoría de periodistas de

otras secciones no lo necesitan” por lo que se puede deducir que la diferencia entre puntuaciones se puede deber a la percepción de la afirmación, unos centrandó la respuesta en periodismo de datos mientras que otros han pensado en los compañeros de otras áreas.

La pregunta “La programación potencia la creación de contenidos periodísticos” es la que mayor nota ha obtenido (8,9). Su mediana es de 9,5 y la nota que más le han puesto es 10. A su vez es la que menos desviación típica tiene (1,6). Por lo que ha sido la pregunta con mayor consenso y puntuación siendo la nota más baja un 5 (Civio).

A la pregunta sobre qué perfiles debería tener un equipo de datos ideal, los entrevistados han contestado mayormente que en todo equipo de datos hace falta la presencia de un periodista que sepa darle un enfoque a los datos y contar bien la historia para transmitirla mejor. Después de la figura del periodista, un perfil de los más mencionados es el del analista de datos. Incluso varios entrevistados contestaron que la presencia de varios de ellos sería clave para el desarrollo de varios temas a la vez o de temas con una gran cantidad de datos que una sola persona tardaría mucho tiempo. Estos analistas deben de saber normalizar los datos y extraerlos.

El siguiente especialista es el de programador. En este apartado se diferencia entre los programadores back-end y front-end. Los programadores back-end se dedican a diseñar la lógica y las soluciones para que todas las acciones solicitadas en una página web sean ejecutadas de manera correcta. Trabaja del lado del servidor y procesa la información recibida a través del front-end. Es responsable de la programación del sitio en todos sus componentes por lo que su trabajo lo ven muy necesario para que la experiencia final del usuario sea la correcta. Por su parte, los front-end se encargan de desarrollar la parte visual de la web. Diseñan la estructura, la tipografía, la colorimetría, imágenes, banners, etc. Ven en estos una especialidad clave para los resultados finales de las visualizaciones de datos, aunque algunos entrevistados piensan que es mejor contar además con un experto en visualización digital para mejorar la calidad artística de las visualizaciones.

Dos de los periodistas consultados añaden que para que estos perfiles encajen bien debe existir la figura de un coordinador que entienda tanto los trabajos periodísticos como los técnicos.

Uno de los entrevistados opina que más que los perfiles se necesitaría cambiar un poco la filosofía del periodismo digital de última hora ya que un equipo de datos necesita “principalmente tiempo, gente y dinero” haciendo referencia a poder pagar programas y herramientas con planes de pago.

4.7 Equipos de datos

Después de su idea ideal de periodismo de datos, les hemos preguntado sobre el equipo de datos que contiene su medio. De estos, los cuatro medios con más audiencia según ComScore (El Confidencial, El País, RTVE y elDiario.es) tres de ellos cuentan con tres profesionales en sus equipos de datos. Dos de esos equipos cuentan solamente con periodistas especializados y con conocimientos en análisis de datos y programación, mientras el otro medio cuenta con un periodista con conocimientos de análisis y programación y con dos analistas de datos. Además, dos de estos equipos cuentan con la colaboración de otras secciones como la sección de “formato” o de “narrativas visuales”. El medio que tiene más de tres trabajadores en su equipo de datos no cuenta con ningún programador ni analista, todos sus empleados son periodistas “con conocimientos en datos”.

Los medios más pequeños tienen equipos más variables. Uno de ellos no tiene un equipo de datos como tal, pero muchos de sus periodistas tienen conocimientos y por ello pueden hacer piezas basadas en datos. Los más pequeños cuenta con dos personas. Uno de ellos cuenta con dos periodistas especializados en datos, pero se dedican a otros ámbitos además de los datos. El otro cuenta con un periodista y un experto en datos y programación y van contratando según su necesidad.

Otro medio cuenta con cuatro periodistas con conocimientos en herramientas y un periodista computacional que hace las tareas de analista y programador en las piezas que es necesario. Otro medio cuenta con solo periodistas y todos tienen conocimientos en programación y tienen especialidades diferentes como análisis o diseño.

El equipo que cuenta con más cantidad de trabajadores tiene 4 periodistas que entienden de datos y además están especializados en un área (economía, salud, política...) y tres técnicos, un informático, un desarrollador web y un diseñador de gráficos interactivos.

Como vemos, en los equipos de datos no se cuenta directamente con los perfiles deseados como los programadores back-end o front-end si no que los programadores tienen que hacer varias tareas. Por otra parte, los analistas en datos no están presentes en todos los medios, pero sí en una gran parte de ellos y vemos cómo los periodistas ocupan el gran porcentaje de los trabajadores de los equipos de datos en España.



5. Conclusiones

Esta investigación se centra en la forma de trabajar de los medios españoles en una disciplina en auge que pese a tener muchas características comunes con otras especialidades más antiguas como el periodismo de precisión, se diferencia de éstas por las nuevas tecnologías con las que se trabaja (grandes bases de datos o programación). Por esta circunstancia novedosa es interesante saber la forma en la que los medios están trabajando esta modalidad y ver las posibles vías de mejora y sus puntos débiles.

Los resultados obtenidos en esta investigación muestran que todavía hay medios que no trabajan con programación ni para analizar los datos ni para visualizarlos. Esto hace que no puedan trabajar con bases de datos que se actualizan durante el tiempo, ni trabajar con grandes cantidades de datos ya que las herramientas utilizadas para estas labores (Excel, Google SpreadSheets, Openrefine...) no soportan archivos tan grandes.

5.1 Programación en periodismo

Los periodistas utilizan la programación en caso de que cuenten con bases de datos muy amplias o de constante actualización, como indicamos en las hipótesis (H2). También recurren a la programación a la hora de hacer web scraping (raspado) para sacar datos que hay en páginas web ya que opinan que las herramientas para esta función son limitadas y muy difíciles de entender su funcionamiento. Pero no pueden realizar piezas trabajadas con programación para los temas de los que no disponen de mucho tiempo y valoran trabajar con herramientas alternativas cuando hay poco volumen de datos “para ahorrar tiempo”.

Entre los lenguajes de programación más utilizados, JavaScript es el lenguaje que está presente en todos los medios que trabajan con programación. Esto se debe a su potente librería de visualización, D3, ya que JS es muy poco usado en las otras fases del proceso. Como veíamos en las hipótesis, R es el lenguaje de programación por excelencia para la obtención, limpieza y análisis seguido por Python que pese a ser el lenguaje mejor

valorado para el periodismo de datos, su uso es menor (H3). Los otros lenguajes mencionados durante la investigación (Ruby, PHP o SQL) tienen un uso muy testimonial. Pero todos afirman que la elección del lenguaje se centra más en la dominación del lenguaje más que en factores de calidad, ya que todos pueden hacer trabajos similares y lo que hace elegirlos es la potencialidad que se le pueda sacar. Por este motivo muchos profesionales ven importante especializarse de un solo lenguaje para poder sacar el máximo rendimiento.

Para trabajar con estos lenguajes utilizan librerías, que son unas implementaciones funcionales que ofrecen una interfaz bien definida para la funcionalidad que se invoca, que les permite ahorrar escribir mucha cantidad de código y simplifica mucho las tareas. Las más utilizadas en R son las que forman el conjunto Tidyverse (H3): dplyr para la manipulación de datos, tidyr para ordenar los datos y readr para leer datos de csv, tsv y fwf. El conjunto contiene más librerías como ggplot2, purrr, tibble, stringr y forcats pero no son tan utilizados como los tres primeros. Otra librería muy recurrida es googlesheets (desde 2021 renombrada a googlesheets4) que permite enlazar tus bases de datos de R a Google SpreadSheets.

En Python la librería más popular es Pandas (H3) para el manejo y análisis de datos. Otras librerías menos populares son BeautifulSoup para el web scraping, Selenium para probar y registrar interacciones en páginas web, Geopandas, para trabajar con datos geoespaciales, Geocoder, librería especializada en geocodificación y NumPy que da soporte para crear vectores y matrices grandes multidimensionales.

Por su parte, JavaScript cuenta con la librería más popular, D3 (H3) en la que realizan las visualizaciones especiales que no pueden llevarse a cabo con las herramientas alternativas. Otras librerías utilizadas son Leaflet que permite realizar mapas interactivos, Vue.js y Node.js con menos usos. Por eso es muy importante que además de dominar uno de los lenguajes de análisis de datos, también se aprenda el uso de esta librería para visualizar los datos de una forma más personalizada y sin depender de terceras empresas.

Ruby es el lenguaje que menos medios utilizan (solo Civio) y las librerías que utilizan son: Nokogiri, que sirve para leer, escribir, modificar y consultar documentos XML y HTML y Mechanize, que se utiliza para automatizar la interacción con sitios web.

5.2 Herramientas alternativas a la programación más utilizadas

El estudio refleja un mayor uso de herramientas alternativas por parte de los periodistas de datos y existe una amplia lista de herramientas con las que se consiguen resultados similares a los obtenidos con programación, pero no permiten tener mucha personalización y esto provoca que muchas visualizaciones publicadas por los medios de comunicación son iguales cambiando simplemente el color. También se depende de las estrategias empresariales que tomen los dueños de estas herramientas ya que pueden variar los servicios y provocar que no puedas seguir utilizándola y tengas que buscar otras herramientas similares.

La herramienta más utilizada para la obtención de datos es Tabula. La utilizan para poder extraer tablas de pdf de una manera muy sencilla. Esta fase del proceso es la que cuenta con más herramientas diferentes (9), pero en lo que coinciden todos es en que la mayor parte de sus bases de datos las han obtenido directamente desde las páginas webs de distintas administraciones.

Openrefine, además de ser la segunda preferida para la obtención, es la que más utilizan para la limpieza de los datos. Una de las funcionalidades que más demandada está es el *clustering* (agrupa valores que están representados de formas diferentes, pero se refieren a la misma. Ejemplo: Av. Miguel Hernández y Avda. Miguel Hernández). Herramientas como Excel (H4), Google SpreadSheet o Calc (Excel de Linux) tienen presencia en la limpieza de datos, pero donde destacan es en el análisis. Se podría decir que Excel es la herramienta clave a la hora de trabajar con bases de datos. Esto se da por la usabilidad y la intuitividad de los programas, al dar la sensación de poder tocar los números. Además, para analizar los datos se suelen utilizar herramientas de visualización (Datawrapper o Flourish) para ver qué patrones dibujan los datos y poder descubrir historias detrás de los datos.

En la visualización la más recurrida es Datawrapper (H4) (herramienta de fácil acceso para hacer visualizaciones interactivas que pueden embeberse y potenciar las narrativas del trabajo periodístico) por encima de Flourish (plataforma para la visualización de datos y la narración de historias) por poco. Esta diferencia se da debido a la rápida carga de la primera y su estilo más limpio. Hay medios que solo utilizan la segunda en caso de que Datawrapper no pueda realizar la visualización requerida. Otras herramientas son Mapbox, que pese a necesitar conocimientos de JS, no llega a ser una librería si no que es un proveedor de mapas web; Qgis que es un Sistema de Información Geográfica (SIG) de software libre y de código abierto que permite realizar mapas estáticos y trabajar con datos geospaciales; Infogram plataforma de visualización de datos; uMap que permite crear mapas con capas de OpenStreetMap y RAWGraphs que permite realizar gráficos de D3 estáticos. También se utilizan para visualizar futuros datos (como los electorales) con datos anteriores y poder realizar un trabajo previo para agilizar el proceso de publicación. De esta forma se prepara una plantilla en la que solamente se deben actualizar las cifras.

Como hemos dicho anteriormente, esto provoca que los medios que trabajan con estas dos herramientas tengan un estilo similar pese a que se puedan editar los colores o las tipografías. Una buena manera de evitar esta uniformidad es crear plantillas propias reutilizables en JS (forma de trabajar de Civio) que están preparadas para que el periodista introduzca los datos de una forma sencilla. Así se unifican diseños y se crea un estilo propio en el medio diferenciado de la competencia.

5.3 Perfil profesional

El profesional de un periodista de datos debe saber preguntar a los datos para sacar historias interesantes. Otra característica importante para trabajar en esta especialidad es no tener miedo a la tecnología, ya que la mayor parte del trabajo requiere de habilidades informáticas. Entender las bases de datos (H5) y tener una mente analista también son unas características muy importantes para los propios periodistas. Ser ordenado también es una cualidad muy relevante para poder realizar estas tareas. Estas características coinciden con las propuestas por otras investigaciones previas.

Por otra parte, no ven necesario que el periodista sea el que controle la programación si no que debe tener unos conocimientos básicos para poder entender el trabajo de los especialistas que trabajan en su medio. Esto no es aplicable a los medios que no cuentan con programadores en sus filas, por lo que esos periodistas deberán aprender más programación para conseguir un producto final sin necesidad de recurrir a la programación. Saber programación aporta una mayor independencia a los profesionales.

Según la perspectiva de los periodistas es factible conseguir un buen nivel de programación si se cuenta con tiempo suficiente para formarse, aunque sea de manera autodidacta. Pero el tiempo no es un recurso con el que cuenten los periodistas y no se puede llegar a formar durante épocas de trabajo. No hay un pensamiento claro sobre si todos los periodistas deberían saber de programación, pero la mayoría piensa que no es necesario que todos sepan porque hay secciones que no necesitan la programación para crear piezas (como la actualidad política, la opinión o la sociedad) pero que puede ser un recurso puntual para algunos temas de esas secciones.

En cuanto a la formación sobre periodismo de datos ofrecida por los grados, muchos no creen que se impartan los suficientes conocimientos para salir de la carrera sabiendo utilizar bases de datos o herramientas necesarias para la labor, pero se ve una tendencia a conseguir esa formación necesaria de cara al futuro. Los másteres especializados están mejor valorados y opinan que además cada vez se va aumentando la formación que ofrecen. Por otra parte, la formación ofrecida por los medios a sus trabajadores está bien valorada por la mitad de la muestra, aunque este trabajo de formación la mayoría de las veces se imparte entre los propios compañeros a través de micro cursos en los que enseñan a los demás miembros de la redacción herramientas o el uso de algún lenguaje de programación. Ofrecer cursos especializados en programación para periodismo agilizaría el proceso de aprendizaje y se podría conseguir un nivel aceptable dedicándole menos tiempo del que se necesita de manera autodidacta.

En esta investigación se ha demostrado que las redacciones no cuentan con los miembros ni los perfiles con los que les gustaría contar a los propios periodistas. La mayoría de los equipos de datos cuentan con periodistas con conocimientos en bases de datos y herramientas, pero escasean de perfiles más técnicos como programadores, analistas o

informáticos. Además, la mayoría de las veces no cuentan ni con el tiempo ni con los recursos que les gustaría para poder realizar piezas más elaboradas. Algunos medios aun no cuentan con una sección de datos al uso, cuentan con redacciones de periodistas especializados en datos. Tener trabajadores sin estar organizados fomenta el trabajo individual sin la creación de proyectos grupales más potentes. Por otra parte, muchos de esos especialistas en datos no ven que la programación sea un recurso muy utilizado a la hora de crear contenidos, aunque la gran mayoría piensen que utilizar lenguajes de programación potencia los contenidos. Esto es un ejemplo claro de la poca confianza de los medios hacia esta tecnología y lo beneficioso que puede llegar a ser según sus empleados.

Pese a las limitaciones del estudio en cuanto a número de medios seleccionados y profesionales entrevistados (10) se puede observar que la mayoría de los medios que trabajan con datos no cuentan con el equipo óptimo para realizar un trabajo potente en lo referente a los datos y muchos de ellos mencionan como ejemplo a seguir la estrategia del NY TIMES en cuanto a recursos y cantidad de especialistas de diferentes áreas (programadores, geógrafos, analistas, matemáticos...). La escasez de recursos con los que cuenta actualmente el periodismo de datos en España trae como consecuencia no poder realizar investigaciones profundas apoyándose de especialistas en estas áreas.

6. Anexos

6.1 Periodistas de datos de los medios de comunicación seleccionados

Medios	Matriz	Periodistas	Cargo
Civio	Digital	David Cabo	Co-Director, Ingeniero
Civio	Digital	Antonio Hernández	Periodista de datos y desarrollador web
Civio	Digital	Eva Belmonte	Co-Directora, Periodista
Civio	Digital	Miguel Ángel Gavilanes	Periodista de datos
Civio	Digital	María Álvarez del Vayo	Periodista de datos
Civio	Digital	Carmen Torrecillas	Programadora web, visualización de datos y diseño gráfico
Datadista	Digital	Antonio Delgado	Cofundador - Periodista de datos
Datadista	Digital	Ana Tudela	Cofundadora - Periodista de investigación
El Confidencial	Digital	Marta Ley	Coordinadora de la sección de datos
El Confidencial	Digital	Carlos Muñoz Díaz	Desarrollador web y periodista de datos
El Confidencial	Digital	Laura Martin	Periodista visual
El Confidencial	Digital	Dario Ojeda	Periodista de datos y reportajes
El Confidencial	Digital	María Zuil	Periodista de datos
El Mundo	Prensa	Gabriela Galarza	Periodista de datos
El Mundo	Prensa	Jose F. Leal	Periodista de investigación
El País	Prensa	Mariano Zafra	Redactor jefe de Narrativas Visuales
El País	Prensa	Daniele Grasso	Periodista de datos
El País	Prensa	Borja Andrino	Analista de datos
El País	Prensa	Kiko Llaneras	Analista de datos
El País	Prensa	Luis Sevillano	Desarrollador gráfico
eldiario.es	Digital	Raúl Sánchez	Coordinador datos y visualización
eldiario.es	Digital	Victòria Oliveres	Periodista de datos
eldiario.es	Digital	Ana Ordaz	Periodista de datos
La sexta	Audiovisual	Verónica Ramírez	Periodista de datos
La sexta	Audiovisual	Dámaso Mondéjar	Periodista digital
Maldita	Digital	Nacho Calle	Coordinador de Maldito Dato
Maldita	Digital	Sergio Sangiao	Coordinador de Maldito Dato
Maldita	Digital	Bea Lara	Coordinadora de maldit@s
Maldita	Digital	Ximena Villagrán	Coordinación de proyectos e impacto
Maldita	Digital	Yuly Jara	Periodista de datos
Maldita	Digital	Jose Molina	Periodista computacional
Maldita	Digital	Adela Vived	Periodista de datos, verificadora
Maldita	Digital	Rubén Díaz	Periodista de datos
Newtral	Digital	Alba Martín Campos	Periodista de datos y visualización
Newtral	Digital	María Blanco	Periodista de datos
Newtral	Digital	Ana Gómez	Periodista de datos
Newtral	Digital	Lorens Leon	Programador
Newtral	Digital	Javier Beltran	Analista de datos
RTVE	Audiovisual	Paula Guisado	Responsable unidad de datos
RTVE	Audiovisual	Cristina Pozo García	Unidad de DatosRTVE
RTVE	Audiovisual	José Ángel Carpio	Unidad de DatosRTVE
RTVE	Audiovisual	Jaime Gutiérrez	Unidad de DatosRTVE (visualizaciones)
RTVE	Audiovisual	Ana Martín Plaza	Unidad de DatosRTVE

Storydata	Servicio comunicación	Eli Vivas	Cofundadora - Periodista de datos
Storydata	Servicio comunicación	Sílvia Galilea	Cofundadora - Periodista de datos
Storydata	Servicio comunicación	Laia Brufau Pla	Periodista de datos
Storydata	Servicio comunicación	Carina Bellver	Cofundadora - Periodista de datos
Storydata	Servicio comunicación	Laura Navarro Soler	Periodista de datos

Fuente: Elaboración propia



6.2 Librerías

6.2.1 Librerías JavaScript

D3: <https://github.com/d3/d3/wiki>

Vue.js: <https://v3.vuejs.org/guide/installation.html>

Leaflet: <https://leafletjs.com/reference-1.7.1.html>

Node.js: <https://nodejs.org/docs/latest-v15.x/api/documentation.html>

6.2.2 Librerías R

tidyverse:

dplyr: <https://dplyr.tidyverse.org/>

tidyr: <https://tidyr.tidyverse.org/>

readr: <https://readr.tidyverse.org/>

ggplot2: <https://ggplot2.tidyverse.org/>

purrr: <https://purrr.tidyverse.org/>

tibble: <https://tibble.tidyverse.org/>

stringr: <https://stringr.tidyverse.org/>

forcats: <https://forcats.tidyverse.org/>

googlesheets4: <https://cran.r-project.org/web/packages/googlesheets4/googlesheets4.pdf>

countrycode: <https://cran.r-project.org/web/packages/countrycode/countrycode.pdf>

rmapshaper: <https://cran.r-project.org/web/packages/rmapshaper/rmapshaper.pdf>

data.table: <https://cran.r-project.org/web/packages/data.table/data.table.pdf>

6.2.3 Librerías Python

Pandas: <https://pandas.pydata.org/docs/pandas.pdf>

beautifulSoup: <https://readthedocs.org/projects/beautiful-soup-4/downloads/pdf/latest/>

selenium: <https://readthedocs.org/projects/selenium-python/downloads/pdf/latest/>

geopandas: <https://geopandas.readthedocs.io/en/latest/docs.html>

geocoder: <https://buildmedia.readthedocs.org/media/pdf/geocoder/latest/geocoder.pdf>

numPy: <https://numpy.org/doc/stable/numpy-user.pdf>

6.2.4 Librerías Ruby

Nokogiri: <https://nokogiri.org/rdoc/index.html>

Mechanize: <https://www.rubydoc.info/gems/mechanize/>



7. Bibliografía

- Arcila-Calderón, C., Barbosa-Caro, E., & Cabezuelo-Lorenzo, F. (2016). Técnicas big data: análisis de textos a gran escala para la investigación científica y periodística. *El Profesional de La Información*, 25(4), 623. <https://doi.org/10.3145/epi.2016.jul.12>
- Barcel, M., Caribe, R., & Sociales, C. (2020). *Competencias informacionales en el contexto del periodismo digital*.
- Codina, L. (2016). Tres dimensiones del periodismo computacional. Intersecciones con las ciencias de la documentación. *Anuario ThinkEPI*, 10, 200. <https://doi.org/10.3145/thinkepi.2016.41>
- Crucianelli, S. (2013). ¿Qué es el periodismo de datos? *Cuadernos de Periodistas*, 26, 106–124.
- Dader, J. L. (2002). *Periodismo de Precisión. Vía socioinformática de descubrir noticias*. 174.
- Del, R. Y. E., & Datos, P. D. E. (2017). *TecCom Studies TecCom Studies*. 6, 4–7.
- Díaz-Campo, J., & Chaparro-Domínguez, M. Á. (2020). Computational journalism and ethics: An analysis of deontological codes of Latin American. *Icono14*, 18(1), 10–32. <https://doi.org/10.7195/RI14.V18I1.1488>
- Domínguez, M. Á. C. (2014). Nuevas formas informativas: el periodismo de datos y su enseñanza en el contexto universitario/New information methods: data journalism and its teaching in the university context. *Historia y Comunicación Social*, 19, 43–54. http://roble.unizar.es/search*spl/i?SEARCH=11370734%5Cnhttp://linksource.ebsco.com/linking.aspx?sid=ProQ%3Aabiglobal&fmt=journal&genre=article&issn=11370734&volume=19&issue=&date=2014-01-01&spage=43&title=Historia+y+Comunicación+Social&atitle=Nuevas+fo
- Ferreras Rodríguez, E. M. (2013). Aproximación teórica al perfil profesional del ‘Periodista de datos.’ *Revista ICONO14. Revista Científica de Comunicación y Tecnologías Emergentes*, 11(2), 115. <https://doi.org/10.7195/ri14.v11i2.573>
- Ferreras Rodríguez, E. M. (2016). El Periodismo de Datos en España. *Estudios Sobre El Mensaje Periodístico*, 22(1), 255–272. https://doi.org/10.5209/rev_ESMP.2016.v22.n1.52594
- Gertrudis Casado, M., Gertrudix Barrio, M., & Álvarez García, S. (2016). Competencias informativas profesionales y datos abiertos: Retos para el empoderamiento ciudadano y el cambio social. *Comunicar*, XXIV(47), 39–47.

- Gran, T. A., Para, E., Arcila-calderón, C., Cabezuelo-lorenzo, E. B. F., Rey, U., & Carlos, J. (2016). *TÉCNICAS BIG DATA : ANÁLISIS DE*.
- Guallar, J. (2013). Introducció al periodisme de dades. Característiques, panoràmica i exemples. *Item: Revista de Biblioteconomia i Documentació*, 0(57), 23–36.
- Hammond, P. (2017). From computer-assisted to data-driven: Journalism and Big Data. *Journalism*, 18(4), 408–424. <https://doi.org/10.1177/1464884915620205>
- Lindgren, C. A. (2021). Writing With Data: A Study of Coding on a Data-Journalism Team. In *Written Communication* (Vol. 38, Issue 1). <https://doi.org/10.1177/0741088320968061>
- López López, P. J. (2018). *La enseñanza-aprendizaje del periodismo de datos en las facultades españolas. Estudios y propuestas de viabilidad*. <https://eprints.ucm.es/50688/1/T40749.pdf>
- Murphy, B. M., & Barton, J. (2014). From a Sea of Data to Actionable Insights : Big Data and What It Means for Lawyers. *Intellectual Property & Technology Law Journal*, 26(3), 8–18.
- Nunan, D., & Di Domenico, M. (2013). Market Research and the Ethics of Big Data. *International Journal of Market Research*, 55(4), 505–520. <https://doi.org/10.2501/ijmr-2013-015>
- Parasie, S., & Dagiral, E. (2013). Data-driven journalism and the public good: “Computer-assisted-reporters” and “programmer-journalists” in Chicago. *New Media and Society*, 15(6), 853–871. <https://doi.org/10.1177/1461444812463345>
- Pérez-Montoro, M. (2017). Visualización de información en 2016: conceptos, contenidos y tecnología. *Anuario ThinkEPI*, 11, 284. <https://doi.org/10.3145/thinkepi.2017.55>
- Polaino, R. C., & Vivar, J. F. (2020). Periodismo de datos y visualización. *Periodismo y Nuevos Medios: Perspectivas y Retos*, October.
- Sánchez Calero, M. L., & Mancinas-Chávez, R. (2017). Periodismo con datos no es igual que Periodismo de datos. *TecCom Studies*, 8, 53–69. https://www.researchgate.net/publication/262871268_El_periodista_frente_a_la_crisis_del_periodismo_en_la_era_digital
- Sánchez García, P., Redondo, M., & Arias, D. (2018). El perfil del periodista de datos en España: formación, fuentes y herramientas. *Doxa Comunicación: Revista Interdisciplinar de Estudios de Comunicación y Ciencias Sociales*, 27, 273–293. <https://dialnet.unirioja.es/servlet/articulo?codigo=6774523>
- Sánchez-García, P., Campos-Domínguez, E., & Berrocal Gonzalo, S. (2015). Las funciones inalterables del periodista ante los perfiles multimedia emergentes. *Revista Latina de Comunicación Social*, 70, 187–208. <https://doi.org/10.4185/RLCS-2015-1042>

Vállez, M., & Codina, L. (2018). Periodismo computacional: evolución, casos y herramientas. *El Profesional de La Información*, 27(4), 759. <https://doi.org/10.3145/epi.2018.jul.05>

Varona, D., Herrero Diz, P., Plaza, J. F., Pulido, C., & Camarero, E. (2016). Elementos clave del Periodismo de datos y la ciudadanía. *Cuadernos Artesanos de Comunicación*, 112, 293–315. <https://doi.org/10.4185/cac112>

