



Universidad Nacional Mayor de San Marcos

Universidad del Perú. Decana de América
Facultad de Letras y Ciencias Humanas
Escuela Profesional de Comunicación Social

El periodismo de datos y la Ley de Transparencia y Acceso a la Información Pública en Perú. Casos Ojo Público y Convoca

TESIS

Para optar el Título Profesional de Licenciado en Comunicación
Social

AUTOR

Rocío Lizbetty ROMERO BENITES

ASESOR

Dra. Carolina Leonor ALBORNOZ FALCÓN

Lima, Perú

2022



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

Referencia bibliográfica

Romero, R. (2022). *El periodismo de datos y la Ley de Transparencia y Acceso a la Información Pública en Perú. Casos Ojo Público y Convoca*. [Tesis de pregrado, Universidad Nacional Mayor de San Marcos, Facultad de Letras y Ciencias Humanas, Escuela Profesional de Comunicación Social]. Repositorio institucional Cybertesis UNMSM.

Metadatos complementarios

Datos de autor	
Nombres y apellidos	Rocío Lizbetty Romero Benites
Tipo de documento de identidad	DNI
Número de documento de identidad	70328925
URL de ORCID	https://orcid.org/0000-0003-2935-1974
Datos de asesor	
Nombres y apellidos	Carolina Leonor Albornoz Falcón
Tipo de documento de identidad	DNI
Número de documento de identidad	06532624
URL de ORCID	https://orcid.org/0000-0001-7366-8210
Datos del jurado	
Presidente del jurado	
Nombres y apellidos	Dennis Javier Vargas Marín
Tipo de documento	DNI
Número de documento de identidad	09178836
Miembro del jurado 1	
Nombres y apellidos	María Mercedes Palomino Gonzales
Tipo de documento	DNI
Número de documento de identidad	43937497
Miembro del jurado 2	
Nombres y apellidos	Carlos Ricardo Gonzales García
Tipo de documento	DNI
Número de documento de identidad	07534446
Datos de investigación	
Línea de investigación	E.3.1.3. Sociedad de la información y tecnologías de la comunicación

Grupo de investigación	Interculturalidad y Comunicación – Incomun
Agencia de financiamiento	Sin financiamiento
Ubicación geográfica de la investigación	Edificio: Facultad de Letras y Ciencias Humanas de la Universidad Nacional Mayor de San Marcos País: Perú Departamento: Lima Provincia: Lima Distrito: Lima Latitud: -12.05682743216602 Longitud: -77.08151739237492
Año o rango de años en que se realizó la investigación	2018 - 2021
URL de disciplinas OCDE	Periodismo https://purl.org/pe-repo/ocde/ford#5.08.01 Medios de comunicación, Comunicación socio-cultural https://purl.org/pe-repo/ocde/ford#5.08.04

Escuela Profesional de Comunicación Social

ACTA DE SUSTENTACIÓN DE TESIS

Reunidos en sesión virtual el día 12 de septiembre del 2022 a las 17:00 horas, el Jurado Evaluador integrado por el Dr. Dennis Javier Vargas Marín (presidente), Mg. María Mercedes Palomino Gonzales (miembro), Mg. Carlos Ricardo Gonzales García (miembro), Dra. Carolina Leonor Albornoz Falcón (asesora), dio inicio a la sustentación de la tesis titulada **El periodismo de datos y la Ley de Transparencia y Acceso a la Información Pública en Perú. Casos Ojo Público y Convoca**, presentado por la bachiller **Rocío Lizbetty Romero Benites**, para optar al Título Profesional de Licenciado en Comunicación Social.

Concluida la sustentación, el jurado procedió a la calificación con el siguiente resultado:

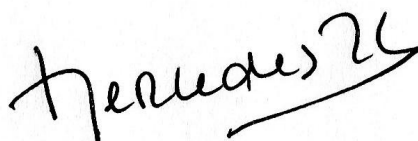
Aprobado con máximos honores (19 - diecinueve)

Luego del proceso de sustentación y la calificación correspondiente, se le comunicó al graduando el resultado obtenido y el Jurado recomienda a la Facultad que se le otorgue el título profesional de Licenciado en Comunicación Social.

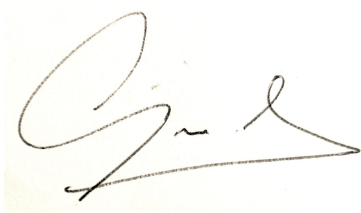
Siendo las 18:00 horas, se concluyó el acto por lo cual, los miembros del jurado dan fe de lo actuado firmando la presente Acta.



Dr. Dennis Javier Vargas Marín
Presidente (Asociado)



Mg. María Mercedes Palomino Gonzales
Miembro (Auxiliar)



Mg. Carlos Ricardo Gonzales García
Miembro (Asociado)



Dra. Carolina Leonor Albornoz Falcón
Asesora (Principal)



Escuela Profesional de Comunicación Social
“Año del Fortalecimiento de la Soberanía Nacional”

INFORME DE EVALUACIÓN DE ORIGINALIDAD

Directora Escuela Profesional de Comunicación Social

Dra. Jacqueline Oyarce Cruz

Operador del programa informático de similitudes

Dra. Jacqueline Oyarce Cruz

Documento evaluado: El periodismo de datos y la Ley de Transparencia y Acceso a la Información Pública en Perú. Casos Ojo Público y Convoca.

(Tesis para optar el Título Profesional de Licenciada en Comunicación Social)

1. Autor del documento: Romero Benites Rocío Lizbety
2. Fecha de recepción de documento 25/08/2022
3. Fecha de aplicación del programa informático de similitudes 25/08/2022
4. Software utilizado: Turnitin
5. Configuración del programa detector de similitudes
 - Excluye textos entrecomillados
 - Excluye bibliografía
 - Excluye cadenas menores a 40 palabras
 - Otro criterio (especificar)
6. Porcentaje de similitudes según programa detector de similitudes: cinco (5%).
7. Fuentes originales de las similitudes encontradas: 5% (Se adjunta PDF)
8. Observaciones: Ninguna
9. Calificación de originalidad
 - Documento cumple criterios de originalidad.
10. Fecha del informe: 02/09/2022



Firmado digitalmente por OYARCE
CRUZ Maria Jacqueline FAU
20148092282 soft
Motivo: Soy el autor del documento
Fecha: 02.09.2022 06:14:42 -05:00

Dra. Jacqueline Oyarce Cruz
Directora de la E. P. de Comunicación Social

Letras mayúsculas del Perú y América

Facultad de Letras y Ciencias Humanas / Universidad Nacional Mayor de San Marcos
Calle Germán Amézaga n. ° 375, Lima 1 - Perú. Ciudad universitaria (puerta 3)
Teléfonos: (051) (01) 452 4641 / (051) (01) 619 7000 - www.lettras.unmsm.edu.pe

A mi mamá Gladys.

Si no existieran los recuerdos, hoy sería un día triste.

Tengo suerte, aunque a veces no es suficiente, de mirarte en fotografías,

reconocer tu silbido en labios extraños

y sentir que aún estás a mi lado.

ÍNDICE

	Pág.
RESUMEN	7
ABSTRACT	8
INTRODUCCIÓN	9
 CAPÍTULO I: PROBLEMA DE INVESTIGACIÓN	
1.1. Descripción del problema	12
1.2. Definición del problema	13
1.2.1. Problema general	13
1.2.2. Problemas específicos	13
1.3. Objetivos	14
1.3.1. Objetivo general	14
1.3.2. Objetivos específicos	14
1.4. Justificación e importancia de la investigación	15
1.5. Delimitación de la investigación	16
 CAPÍTULO II: REVISIÓN DE LITERATURA	
2.1. Antecedentes del estudio	17
2.2. Bases teóricas	21
2.2.1. El periodismo de datos	21
2.2.1.1. Antecedentes del periodismo de datos	28
2.2.1.2. Ámbitos del periodismo de datos	31
2.2.1.3. Metodología del periodismo de datos	32
2.2.2. La Ley de Transparencia y Acceso a la Información Pública	39
2.2.2.1. Denegatoria de información	41
2.2.2.2. El recurso de apelación	43
2.2.2.3. El hábeas data	44
2.2.2.4. El periodismo de datos y el acceso a la información pública	45
2.2.3. Los medios nativos digitales peruanos <i>Ojo Público</i> y <i>Convoca</i>	46
2.2.3.1. El proyecto periodístico <i>Excesos sin castigo</i>	49
2.2.3.2. El proyecto periodístico <i>Memoria robada</i>	51

CAPÍTULO III: HIPÓTESIS Y VARIABLES

3.1.	Hipótesis	53
3.2.	Variables	53
3.3.	Operacionalización de variables	55

CAPÍTULO IV: MATERIALES Y MÉTODOS

4.1.	Diseño de la investigación	56
4.2.	Población y muestra	56
4.3.	Técnicas de recolección de datos	59
4.3.1.	Encuesta	59
4.3.2.	Entrevista	60

CAPÍTULO V: RESULTADOS

5.1.	Presentación y análisis de resultados	61
5.1.1.	Recopilación de datos	62
5.1.2.	Limpieza de datos	66
5.1.3.	Contextualización y combinación de datos	70
5.1.4.	Análisis de datos	71
5.1.5.	Verificación de datos	72
5.1.6.	Comunicación de datos	74

DISCUSIÓN	77
------------------	----

CONCLUSIONES	80
---------------------	----

RECOMENDACIONES	83
------------------------	----

REFERENCIAS BIBLIOGRÁFICAS	84
-----------------------------------	----

ANEXOS	88
---------------	----

Anexo 1	89
---------	----

Anexo 2	95
---------	----

Anexo 3	99
---------	----

Anexo 4	106
---------	-----

Anexo 5	112
---------	-----

ÍNDICE DE FIGURAS Y TABLAS

	Pág.
Figura 1. Fórmula del periodismo de datos	23
Figura 2. Mapa contextual para el desarrollo del periodismo de datos	23
Figura 3. Las tres dimensiones del periodismo de datos	25
Figura 4. Pirámide invertida del periodismo de datos	33
Figura 5. Modelo de solicitud de acceso a la información pública	40
Tabla 1. Población y muestra	58
Figura 6. Adquisición de conocimientos de los periodistas sobre acceso a la información pública	63
Figura 7. Periodistas que presentaron solicitudes de acceso a la información pública	65
Figura 8. Trámites vinculados al proceso de acceso a la información pública conocidos por los periodistas y sus niveles de habilidad antes de iniciar el proyecto	66
Figura 9. Periodistas que procesaron información	67
Figura 10. Programas empleados por los periodistas y sus niveles de habilidad en la fase de limpieza de datos	68
Figura 11. Programas y lenguajes empleados por los tecnólogos y su nivel de habilidad en la fase de limpieza de datos	69
Figura 12. Participación de periodistas en la definición de las estructuras de las bases de datos	70
Figura 13. Programas empleados por los periodistas y sus niveles de habilidad en la fase de análisis de datos	72

Figura 14. Periodistas que participaron en la verificación de datos	73
Figura 15. Tecnólogos que participaron en la verificación de datos	74
Figura 16. Programas y lenguajes empleados por los tecnólogos en el diseño e implementación de los sitios web	76
Figura 17. Programas y lenguajes empleados por los programadores en el diseño e implementación de visualizaciones	76

RESUMEN

Esta investigación describe y explica la aplicación de la metodología del periodismo de datos en investigaciones periodísticas basadas en el análisis de datos, de dos medios nativos digitales independientes de Perú, que tienen como componente principal información obtenida a través de la Ley de Transparencia y Acceso a la Información Pública. Para este propósito, se analizó —a partir de datos recopilados por medio de encuestas y entrevistas— el trabajo realizado por los periodistas y tecnólogos en la ejecución de los proyectos *Memoria robada* y *Excesos sin castigo*, de *Ojo Público* y *Convoca*, respectivamente. Ambas investigaciones periodísticas forman parte de los primeros trabajos de largo plazo publicados por los medios de comunicación en mención, en donde se realizaron cientos de solicitudes de información a instituciones públicas y cuyos datos fueron liberados en sus respectivas plataformas digitales.

Los resultados muestran que la mayoría de los periodistas que participó en la elaboración de los proyectos *Memoria robada* y *Excesos sin castigo* tenía conocimientos de nivel básico e intermedio en el uso de herramientas tecnológicas que facilitaban la labor periodística, sobre todo en las fases de análisis de datos y comunicación de la información con visualizaciones interactivas.

Palabras clave: *Periodismo de datos, Metodología, Ley de Transparencia y Acceso a la Información pública, Memoria robada, Excesos sin castigo.*

Línea de investigación

E.3. Letras y Ciencias Humanas, E.3.1. Humanidades interdisciplinarias, E.3.1.3. Sociedad de la información y tecnologías de la comunicación.

ABSTRACT

This research describes and explains the application of the data journalism methodology in journalistic publications, of two independent digital native media from Peru, that have as a main component data obtained through the Law of Transparency and Access to Public Information. For this purpose, it was analyzed —based on data collected through surveys and interviews— the work done by journalists and technologists in the execution of the projects *Memoria robada* (Stolen Memory) and *Excesos sin castigo* (Unpunished Excesses), of *Ojo Público* and *Convoca*, respectively. Both journalistic investigations are part of the first long-term works published by the media in question, where hundreds of requests for information were made to public institutions and whose data was released on their respective digital platforms.

The results show that most of the journalists who participated in the Stolen Memory and Unpunished Excesses projects had basic and intermediate level knowledge in the use of technological tools, especially in the phases of data analysis and communication of information with interactive visualizations.

Keywords: *Data Journalism, Methodology, Law of Transparency and Access to Public Information, Stolen Memory, Unpunished Excesses.*

INTRODUCCIÓN

El periodismo de datos es una disciplina que en los últimos años ha alcanzado mayor trascendencia en Perú debido a la publicación de investigaciones, basadas en el análisis de datos, que revelan hechos de corrupción y que permiten fiscalizar a las entidades y autoridades públicas. Diversos estudios explican el proceso de trabajo en el periodismo de datos y su desarrollo en las salas de redacción; sin embargo, la mayoría de estos estudios proviene del extranjero.

En este contexto, resulta necesario mencionar que para la ejecución del periodismo de datos se requieren grandes volúmenes de información que pueden ser obtenidos por diferentes vías, como peticiones a entidades públicas. En el caso peruano, ello se logra gracias a la Ley de Transparencia y Acceso a la Información Pública, en adelante Ley, que está vigente desde el 2003. La norma, pese a las limitaciones que posee, permite que cualquier persona acceda a información construida por las entidades de la administración pública o que se encuentra en posesión de ellas sin importar el soporte en el que esté. No obstante, la práctica ha demostrado que las instituciones públicas se rehúsan a entregar información en algunos casos.

Destacan así la labor periodística de dos medios nativos digitales pioneros en la materia: *Ojo Público* y *Convoca*, que realizan investigaciones basadas en el análisis de datos, de corte individual o en colaboración con otros medios de comunicación nacionales o internacionales. Ambas organizaciones se caracterizan por ser independientes y subsistir, fundamentalmente, por los fondos recibidos de instituciones que apuestan por este tipo de periodismo. También, por las recaudaciones que logran de los talleres educativos dictados, la publicación de material bibliográfico, donaciones de lectores, membresías o la venta de *merchandising*.

Para la presente investigación se seleccionaron los proyectos periodísticos *Memoria robada*, de *Ojo Público*, y *Excesos sin castigo*, de *Convoca*, con la finalidad de examinar la metodología de trabajo empleada por sus periodistas y tecnólogos, aquellas personas con conocimientos en tecnología como programadores, desarrolladores, etc. Ambos proyectos se encuentran entre las primeras investigaciones que publicaron estos medios nativos digitales y que requirieron un trabajo considerable de solicitudes de acceso a la información pública a través de la Ley.

Esta investigación, como se detalla más adelante, es exploratoria debido a que abarca un tema poco estudiado en Perú; y descriptiva ya que evalúa la metodología empleada en el trabajo periodístico con datos. Además, es de carácter mixta porque en la etapa de recopilación de información se emplearon técnicas cuantitativas y cualitativas como encuestas y entrevistas que fueron aplicadas a periodistas y tecnólogos de ambos medios digitales, estrategias que permitieron describir el trabajo que se ejecutó detrás de cada proyecto periodístico.

Así, para el desarrollo de este estudio se encuestaron a todos los profesionales que participaron a tiempo completo en las investigaciones en mención. En total fueron nueve periodistas: cinco de *Convoca* y cuatro de *Ojo Público*; y cuatro tecnólogos: tres de *Convoca* y uno de *Ojo Público*. Asimismo, para profundizar y complementar información en varios aspectos, se entrevistaron a tres periodistas que cumplían las funciones de editores y coordinadores de proyectos: uno de *Convoca* y dos de *Ojo Público*.

Los resultados de esta investigación permiten explicar a detalle la labor realizada en cada fase de trabajo del periodismo de datos, las herramientas o programas tecnológicos empleados, las dificultades en torno al acceso a información pública, entre otros detalles mencionados en las páginas posteriores.

En el primer capítulo se explica el problema de investigación y se definen los objetivos, la justificación y delimitación del estudio. En el siguiente apartado se hace referencia a los antecedentes de la presente investigación, al periodismo de datos, la Ley, y a los medios nativos digitales que desarrollaron los proyectos periodísticos *Memoria robada* y *Excesos sin castigo*.

En el capítulo número tres se presentan la hipótesis, variables y la operacionalización de variables; mientras que en el cuarto capítulo se exponen el diseño de la investigación, la población y muestra, y las técnicas de recolección de datos utilizadas.

En el capítulo cinco se presentan y analizan los resultados, considerando las siete fases de trabajo del periodismo de datos: recopilación, limpieza, contextualización, combinación, análisis, verificación y comunicación de información. Luego se expone la discusión, conclusiones y recomendaciones.

El presente estudio tiene como propósito servir de referencia a futuras investigaciones académicas y difundir características o habilidades que se deben fomentar en los próximos periodistas de datos. Asimismo, impulsar el uso de la Ley no solo en periodistas, sino en los demás ciudadanos que opten por fiscalizar el trabajo de las organizaciones públicas o liberar información para que esta esté al alcance de la mayor cantidad de personas y les permita tomar mejores decisiones en su día a día.

CAPÍTULO I: PROBLEMA DE INVESTIGACIÓN

1.1. Descripción del problema

Ojo Público y *Convoca* son dos medios de comunicación nativos digitales pioneros en el periodismo de datos peruano que realizan investigaciones periodísticas centradas en temas como corrupción, derechos humanos, medio ambiente y salud. Para ello, principalmente en sus publicaciones de largo plazo, acceden a datos por distintas vías. La más empleada es la Ley de Transparencia y Acceso a la Información Pública.

Es así que esta normativa, que entró en vigencia el 2003 durante el gobierno del expresidente Alejandro Toledo, se ha convertido en una herramienta para acceder a información de interés ciudadano que permite fiscalizar a las entidades de la administración pública y también a funcionarios.

Entre las investigaciones elaboradas con información obtenida por medio de la citada Ley se encuentran *Memoria robada*, de *Ojo Público*, que abordó el tráfico de patrimonio cultural en América Latina; y *Excesos sin castigo*, de *Convoca*, que expuso el comportamiento de las industrias extractivas en Perú. Para ambos proyectos periodísticos se realizaron decenas de solicitudes de acceso a la información pública —en comparación con otras investigaciones realizadas por estos dos medios desde su fundación hasta el momento en que se inició el presente trabajo de investigación— y se requirió el trabajo de periodistas y tecnólogos durante más de seis meses. De ahí que ambos proyectos sean elegidos para analizar la aplicación de la metodología del periodismo de datos en su elaboración.

No obstante, como se detalla en páginas posteriores, los periodistas y tecnólogos de los dos medios nativos digitales en mención también aplicaron en el desarrollo de los proyectos periodísticos el “ensayo-error”, debido a su poca experiencia en la ejecución de

investigaciones con datos masivos. Es por eso que surge la necesidad de conocer a profundidad cómo los periodistas y tecnólogos involucrados en la elaboración de los proyectos señalados aplicaron la metodología del periodismo de datos, así como sus niveles de conocimiento al respecto. Esto, con la finalidad de que sirva como insumo para otros profesionales y también como referencia para las universidades al momento de armar sus mallas curriculares en las escuelas de periodismo y afines.

1.2. Definición del problema

1.2.1. Problema general

¿Cómo aplicaron la metodología del periodismo de datos los periodistas y tecnólogos de los medios nativos digitales peruanos *Ojo Público* y *Convoca* en la ejecución de las investigaciones —basadas en el análisis de datos— *Memoria robada* y *Excesos sin castigo*, respectivamente, que tienen como fuente principal información obtenida a través de la Ley de Transparencia y Acceso a la Información Pública?

1.2.2. Problemas específicos

- ¿Cuáles son los procesos de recopilación de información que siguieron los periodistas y tecnólogos de *Ojo Público* y *Convoca* en ambas investigaciones periodísticas?
- ¿Cómo fue el proceso de limpieza de información que se realizó en dichas investigaciones periodísticas?
- ¿De qué manera se contextualizaron los datos en los proyectos periodísticos *Memoria robada* y *Excesos sin castigo*?

- ¿Cuáles son los procesos de combinación de datos que siguieron los periodistas y tecnólogos en las investigaciones periodísticas?
- ¿Qué procesos de análisis de datos se han seguido en ambos proyectos de investigación periodística?
- ¿De qué manera se verificaron los datos en los proyectos *Memoria robada* y *Excesos sin castigo*?
- ¿Cómo se comunicaron los resultados de ambas investigaciones basadas en el uso de la Ley de Transparencia y Acceso a la Información Pública?

1.3. Objetivos

1.3.1. Objetivo general

Describir y explicar la metodología del periodismo de datos que aplicaron los periodistas y tecnólogos de los medios nativos digitales peruanos *Ojo Público* y *Convoca* en la ejecución de investigaciones —basadas en el análisis de datos— *Memoria robada* y *Excesos sin castigo*, respectivamente, que tienen como fuente principal información obtenida a través de la Ley de Transparencia y Acceso a la Información Pública.

1.3.2. Objetivos específicos

- Explicar los procesos de recopilación de información que siguieron los periodistas y tecnólogos de *Ojo Público* y *Convoca* en ambas investigaciones periodísticas.
- Detallar el proceso de limpieza de información que se realizó en dichas investigaciones periodísticas.

- Conocer de qué manera se contextualizaron los datos en los proyectos periodísticos *Memoria robada* y *Excesos sin castigo*.
- Explicar los procesos de combinación de datos que siguieron los periodistas y tecnólogos en las investigaciones periodísticas.
- Conocer los procesos de análisis que se han seguido en ambos proyectos de investigación periodística.
- Explicar cómo se verificaron los datos en los proyectos *Memoria robada* y *Excesos sin castigo*.
- Describir la forma en que se comunicaron los resultados de ambas investigaciones basadas en el uso de la Ley de Transparencia y Acceso a la Información Pública.

1.4. Justificación e importancia de la investigación

La presente investigación es útil como base teórica para los estudiantes, periodistas y demás ciudadanos que deseen aprender o fortalecer sus conocimientos sobre la metodología empleada para realizar proyectos periodísticos basados, principalmente, en datos obtenidos por medio de la Ley. Asimismo, permite profundizar, desde el punto de vista práctico y normativo, en la importancia de usar la legislación no solo para obtener datos, sino para convertirlos en información más amigable para consumo de los lectores.

El estudio, además, ayuda a comprender cómo los medios nativos digitales desarrollaron investigaciones con datos y documentos que puede obtener cualquier ciudadano, ejerciendo su derecho de acceso a la información, y cómo responder ante la negativa de las instituciones de otorgar datos públicos.

La autora de esta tesis fue parte del equipo periodístico que desarrolló el proyecto *Memoria robada*, de *Ojo Público*. Esa fue una de las motivaciones que también surgió para

describir y explicar el trabajo metodológico del periodismo de datos empleado por los profesionales involucrados directamente en la ejecución de dichas investigaciones periodísticas. Sin embargo, consideró necesario no incluirse como muestra del estudio para mantener la objetividad de la investigación.

1.5. Delimitación de la investigación

Los medios nativos digitales *Ojo Público* y *Convoca* publican constantemente investigaciones de corto, mediano y largo plazo en las que emplean, además de otras fuentes, información recopilada a través de solicitudes de acceso a la información pública.

Con la finalidad de examinar la aplicación del proceso metodológico del periodismo de datos en proyectos en los que la información se obtuvo, principalmente, a través de la Ley, se seleccionaron dos investigaciones que cumplen con estas características: *Memoria robada* en el caso de *Ojo Público* y *Excesos sin castigo*, de *Convoca*. Estos proyectos fueron desarrollados por un número reducido de periodistas y tecnólogos que trabajaron a tiempo completo en su realización, y como tal se optó por incluirlos a todos ellos como población de estudio para obtener la información pertinente.

CAPÍTULO II: REVISIÓN DE LITERATURA

2.1. Antecedentes del estudio

Con el avance del periodismo de datos se han incrementado las publicaciones académicas que explican esta disciplina: sus características, metodología de trabajo, perfil profesional de los equipos de periodistas y tecnólogos, herramientas digitales, *softwares* y lenguajes de programación empleados en las diferentes fases de trabajo, entre otros aspectos relevantes. Sin embargo, en Perú la producción de investigaciones que abordan este campo aún es baja. Mucho más si se le vincula con la Ley.

Entre los estudios que anteceden a esta investigación, se encuentra el *Periodismo de datos. Antecedentes, ámbitos de desarrollo y aplicación en medios generalistas* de Bueno (2014) en donde se define al periodismo de datos como una disciplina en evolución y se describe la rutina periodística de los medios de comunicación españoles *El Confidencial* y *El Periódico*. La autora observó el trabajo rutinario de ambas empresas periodísticas con la finalidad de comprender la estructura del grupo de trabajo, los conocimientos y habilidades de los periodistas, el proceso de producción de la información, basado en el análisis de datos, así como el acceso a las fuentes y la forma de comunicar los datos. Tras la aplicación de este método cualitativo, se concluyó que “el periodismo de datos da lugar al nacimiento de un nuevo perfil profesional: el periodista de datos”, que tiene habilidades particulares y distantes a las de un periodista tradicional; y que “las tecnologías digitales han modificado las rutinas de producción de la información afectando así a todas las fases del proceso del periodismo de datos” (Bueno, 2014, pp. 36-37). Si bien esta investigación no se centró en explicar la aplicación de la metodología del periodismo de datos en base a trabajos realizados con información pública, es un referente que sentó bases sobre cómo analizar las rutinas periodísticas en dos redacciones diferentes.

Otro de los antecedentes para el presente estudio es *La relación del periodismo de datos con Big Data y Open Data* de Mariño (2012) que planteó como objetivo principal presentar una introducción del periodismo de datos y su vínculo con el *Big Data* y *Open Data*. Una de las conclusiones de la investigación es el hecho de que se requieren diversas habilidades para ejercer el periodismo de datos, por lo que resulta importante trabajar con equipos multidisciplinarios. No obstante, ello aún no se observa con determinación en las salas de redacción ni mucho menos se plantea como posibilidad en la formación académica de los estudiantes de periodismo. En este sentido, también se explicó que dentro de la labor periodística no solo ha sido importante el uso de datos masivos, sino “la concientización sobre liberar información para uso y desuso de los usuarios, sean estos periodistas o no” (Mariño, 2012, p. 56).

Además de estas dos investigaciones, resulta relevante el trabajo *Los primeros pasos del periodismo de bases de datos en Colombia* de Forero (2013) que dentro de sus objetivos se propuso “identificar el grado de reconocimiento y uso del periodismo de bases de datos en medios de comunicación impresos y digitales en Colombia” (p. 10) como lo son *El Espectador* y *La Silla Vacía*, respectivamente. En este estudio se aplicó un enfoque cualitativo y técnicas de investigación como entrevistas, análisis de contenido y encuestas. Y se determinó que la iniciativa de aplicar el análisis de datos en las publicaciones periodísticas parte de los propios reporteros y no de los medios de comunicación en mención. Los periodistas, además, requirieron capacitaciones en técnicas de periodismo de bases de datos para desarrollar sus labores y no se encontraron muy organizados para el cumplimiento de sus funciones. De acuerdo a Forero (2013), los reporteros accedieron a los datos por medio de Internet y peticiones de acceso a la información pública, e incursionaron en el uso de herramientas más avanzadas para la construcción y cruce de información. Sin embargo, se toparon con problemas como la ausencia de información en

formatos amigables que permitan su reutilización inmediata. Esta situación complicó y retrasó la labor periodística en los medios analizados.

Asimismo, como antecedente también se encuentra *El periodismo de datos en 2015* de Tuñas (2015) que tuvo como objetivo principal “identificar aquellos aspectos propios y característicos del periodismo de datos en los diferentes ámbitos convergentes del periodismo: el empresarial, el profesional, el de contenido y el tecnológico” (p. 10). Como proceso metodológico, la investigadora seleccionó cinco trabajos de cada uno de los medios que forman parte de su investigación: *El Mundo* y *The Guardian*, y luego fueron distribuidos por áreas temáticas para revisar sus títulos, tipos de gráficos empleados y cómo se abordó el tema periodístico. Uno de los hallazgos del estudio muestra que las bases de datos públicas no son perfectas y que los gobiernos aún tienen mucho por trabajar para brindar información que la ciudadanía solicita y necesita, y sobre todo hacerlo con transparencia.

En relación al tema de transparencia y acceso a la información pública, destaca la investigación *El acceso a datos como parte del derecho de acceso a la información pública* de Quispe (2016), en donde se definió el marco legal del acceso a la información pública en Perú y se explicó qué sucede cuando un ciudadano no recibe información completa o su petición es negada, pese a haber cumplido los requisitos estipulados en la Ley. De acuerdo a Quispe (2016), el ciudadano puede iniciar un proceso de hábeas data que “podría durar entre 6 y 12 meses en promedio en el Juzgado, 11 meses y medio en la Corte Superior y hasta 25 meses en el Tribunal Constitucional” (p. 73).

Para la presente investigación también se revisó la tesis *Deficiencias y limitaciones normativas al derecho de acceso a la información pública en el Perú* de Cárdenas (2020) que planteó “verificar y comprobar que las deficiencias y limitaciones normativas relacionadas con la falta de independencia de la Autoridad Nacional de Transparencia y

Acceso a la Información Pública afectan el derecho de acceso a la información pública” (p. 8). Para ello, realizó entrevistas semiestructuradas y analizó incumplimientos en materia de acceso a la información pública entre el 2013 y 2018. Una de las conclusiones es la siguiente:

Las deficiencias y limitaciones en el procedimiento para la denegatoria de información establecido por la Ley N° 27806 afecta el derecho de acceso a la información pública toda vez que en la presente tesis se ha demostrado que existen contradicciones entre el TUO de la Ley N° 27806 y el TUO de la Ley N° 27444 respecto a los plazos para interponer un recurso de apelación, que muchas solicitudes de accesos a la información son denegadas alegando una errónea interpretación de las excepciones contempladas por norma y su no ubicación debido a la desorganización de los archivos de la administración pública los que se encuentran en una situación precaria generando un impacto negativo en el ejercicio de este derecho fundamental. (Cárdenas, 2020, p. 300)

Ello es importante para la presente investigación debido a que evidencia la denegatoria de solicitudes para acceder a información creada o en posesión de las entidades públicas, por ejemplo, debido a interpretaciones equivocadas de la Ley. Además, Cárdenas (2020) explica las excepciones a la entrega de información pública, como es la información secreta, reservada y confidencial.

Por último, no se puede dejar de mencionar la investigación de Profeta (2018) titulada *El periodismo de datos como oportunidad para renovar el periodismo*, en donde se analizaron publicaciones de veinte medios de América Latina y entrevistaron a veintidós profesionales del rubro periodístico. En este contexto, se llegó a la conclusión de que el periodismo de datos permite contar historias que solo se pueden encontrar a través de su metodología de trabajo, con nuevas maneras de mostrar los datos para que el lector

adquiera un rol más activo, y promover la fiscalización del poder por medio del acceso a datos públicos y su apertura para los ciudadanos.

2.2. Bases teóricas

2.2.1. *El periodismo de datos*

El periodismo de datos, de acuerdo a Crucianelli (2013b), incluye técnicas de varias disciplinas del periodismo: de investigación, profundidad, analítico, de precisión y asistido por computadoras. Estas dos últimas disciplinas son consideradas antecedentes del periodismo de datos y se detallan en el siguiente apartado. A ello, se suma el uso de grandes cantidades de datos, el empleo de la visualización interactiva y la incorporación de un programador al equipo de trabajo (pp. 106-107). A continuación, se explica brevemente el significado del periodismo de investigación, de profundidad y el analítico.

Janis Karklins, exsubdirector general de Comunicación e Información de la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura, define al periodismo de investigación como:

La tarea de revelar cuestiones encubiertas de manera deliberada, por alguien en una posición de poder, o de manera accidental, detrás de una masa caótica de datos y circunstancias, y en el posterior análisis y exposición pública de todos los datos relevantes. Así, el periodismo de investigación contribuye de manera fundamental a la libertad de expresión y de información. (Lee, 2014, p. 88)

En esta línea, el periodismo de investigación desarrolla “un tema de relevancia social” y “de interés público” (Proética, 2009, p. 9). Además, como menciona el periodista venezolano Ewald Scharfenberg, el reportero acude a un lugar al que no ha sido invitado a buscar historias y hasta a “incomodar”, situación que dista, por ejemplo, de los periodistas que cubren el día a día y que asisten, frecuentemente, a “una rueda de prensa, un evento

deportivo o una boda de famosos” (Instituto Prensa y Sociedad de Venezuela, 2010, p. 7). En definitiva, para el desarrollo del periodismo de investigación es necesario revelar información que está oculta, pero que es de interés público.

En este contexto, es importante mencionar a David Kaplan, quien fue director de la Red Global de Periodismo de Investigación, y manifiesta cuatro características que no definen al periodismo de investigación:

A) No es periodismo de filtración: no se trata de conseguir un documento filtrado por una fuente y reportar sobre este el mismo día. B) No es solamente periodismo especializado: si bien los reporteros especializados usan técnicas de investigación, no necesariamente un periodista experto en una temática es necesariamente un periodista investigativo. C) No es periodismo crítico: el periodismo de investigación puede tener elementos críticos, pero escribir algo que es fuerte y crítico no significa que se cuenta con la información requerida por el periodismo de investigación. D) No es periodismo sobre el crimen y la corrupción: definir al periodismo de investigación como la cobertura de la delincuencia y la corrupción limita el alcance de esta disciplina, si bien existe una superposición entre los dos. (Salamanca, Sierra & Huertas, 2018, p. 22)

Por otro lado, en el periodismo en profundidad, como lo explica Crucianelli (2013b), nadie intenta mantener en secreto lo que se investiga, por lo que se suele obtener información que está al alcance ciudadano como estudios académicos, publicaciones institucionales o censos. En tanto, el periodismo analítico examina una situación compleja, “combina algunos aspectos del periodismo de investigación y la información explicativa” y también “contextualiza un tema de fondo mediante la descripción de detalles históricos y datos estadísticos” (p. 107-108). En resumen, como se aprecia en la Figura 1, Crucianelli (2013b) define al periodismo de datos como la suma de varios aspectos.

Figura 1

Fórmula del periodismo de datos

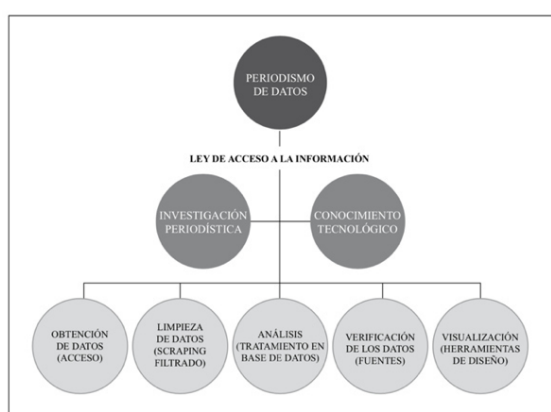
Periodismo de investigación + periodismo de profundidad + periodismo de precisión + periodismo analítico + periodismo asistido por computadora + volumen de datos + visualización interactiva + programación = **periodismo de datos**

Nota. Adaptado de “¿Qué es el periodismo de datos?” (p. 107), por S. Crucianelli, 2013, *Cuadernos de periodistas*, 26.

Flores y Salinas (2013) también han escrito artículos académicos sobre el periodismo de datos y, de acuerdo a ellos, esta especialización se basa en la investigación periodística, el conocimiento tecnológico y en una buena ley de acceso a la información pública: “Sin una ley de transparencia o acceso a la información pública, difícilmente se puede llevar a cabo el desarrollo del periodismo de datos en toda regla” (p. 19). Respecto al conocimiento tecnológico, mencionan que es necesario saber cómo obtener los datos, limpiarlos, analizarlos, verificarlos y visualizarlos, tal como se aprecia en la Figura 2.

Figura 2

Mapa contextual para el desarrollo del periodismo de datos



Nota. Adaptado de “El periodismo de datos como especialización de las organizaciones de noticias en Internet” (p. 19), por J. Flores y C. Salinas, 2013, *Correspondencias & Análisis*, 3.

También, el periodismo de datos se desarrolla gracias a dos aspectos:

La evolución tecnológica, que ha dado lugar a la digitalización de los hechos que suceden en el mundo y el uso de herramientas de intercambio de conocimientos unida a la conciencia cívica; y la necesidad de transparencia de los periodistas, activistas, administradores y personas que encuentran representación en las filas de la cultura informal del *Open Data*. (Flores & Salinas, 2013, p. 20)

En esta línea, Ramonet (2011) señala que el periodismo de bases de datos “es un gran logro de la democracia moderna porque permite buscar y acceder de forma inmediata a bases de información creadas por instituciones públicas o privadas” (p. 71). Asimismo, recalca que el periodista puede hallar información importante que ha sido ocultada intencionalmente o que ha pasado desapercibida. Por su parte, García y Catalina (2018) manifiestan que el periodismo de datos nacería de la relación entre el *Open Data* y el *Big Data*. El primer aspecto tiene que ver con las maneras de recolectar información para luego usarla, reutilizarla y publicarla; mientras que el segundo se vincula con las técnicas, procedimientos y herramientas para trabajar con grandes cantidades de datos.

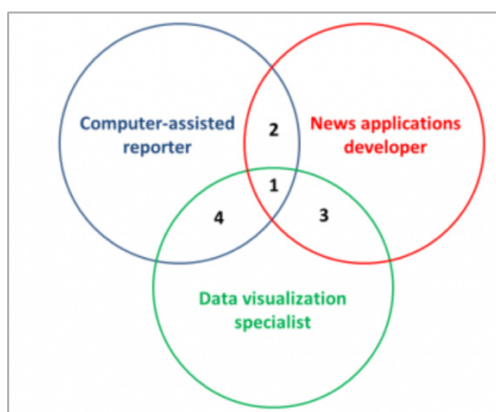
Una de las bases del periodismo de datos es el uso de información abundante. Esta puede estar disponible en la red, en publicaciones de los organismos del Estado, organizaciones no gubernamentales, empresas privadas, etc.; o se puede llegar a ella mediante filtraciones o peticiones de información, dependiendo la documentación o archivos que se requieran y, si la institución a la que se le pide la información, la tiene y puede entregarla. Por ello, Ferreras (2013) explica que el periodismo de datos consiste fundamentalmente en buscar información, seleccionarla y procesarla para convertirla en piezas informativas de fácil comprensión para los ciudadanos (p. 117).

En relación al equipo de trabajo, Gordon (2013) explica la necesidad de que esté conformado por personas con tres habilidades diferentes, y no necesariamente cada una de

ellas debe poseer todas las habilidades a la vez: el periodista asistido por computadora (*computer assisted reporter*), el desarrollador de aplicaciones de noticias (*news applications developer*) y el especialista en visualizaciones de datos (*data visualization specialist*), tal como se aprecia en la Figura 3. De acuerdo a Blanco, Teruel y Martín (2016), “las zonas numeradas aluden a la posibilidad de reunir en una misma persona más de una de estas habilidades, una dificultad que puede disminuir mediante una formación adecuada” (p. 21).

Figura 3

Las tres dimensiones del periodismo de datos



Nota. Adaptado de *Want to build a data journalism team? You'll need these three people*, por R. Gordon, 2013, The Northwestern University Knight Lab (<https://knightlab.northwestern.edu/2013/06/28/want-to-build-a-data-journalism-team-youll-need-these-three-people/>).

El primer profesional al que se hace referencia es al reportero asistido por computadora, quien inicia su trayectoria profesional, usualmente, como reportero y en el camino va adquiriendo experiencia en obtener, analizar y encontrar historias de interés en los datos. Emplea hojas de cálculo y algunos *softwares* para su labor y, también adquiere conocimientos de programación.

El segundo profesional es el desarrollador de aplicaciones de noticias que suele ser un reportero que aprendió de programación o un ingeniero interesado en la práctica periodística. Entre sus habilidades, de acuerdo a Gordon (2013), destacan la programación *Back-End*, que conecta las bases de datos con el servidor en una página web; la configuración y administración de bases de datos; el Lenguaje de Marcado de Hipertexto (*HyperText Markup Language - HTML*), a través del cual se da la estructura a un sitio web; Hojas de Estilo en Cascada (*Cascading Style Sheets - CSS*), que permite estilizar elementos de *HTML*; el lenguaje de programación *JavaScript*, entre otros.

Por último, el tercer profesional que se menciona es el periodista especializado en visualización de datos, quien presenta la información de manera entendible y accesible para los usuarios, principalmente a través de gráficos interactivos. Aquí los lectores, dependiendo de la visualización, pueden navegar, dar clic, comparar o incluso descargar datos. Según Gordon (2013), aquí calza el perfil del periodista que maneja programación *Front-End*, que viene a ser toda la parte visible para los lectores de una página web o aplicaciones, y emplea bibliotecas de código abierto, como *D3*, en donde se aplica el lenguaje de programación *JavaScript*.

En tanto, el profesor Philip Meyer menciona que cuando había escasa información los esfuerzos estaban centrados en cazar y recolectar; mientras que ahora, debido a la abundancia de la información, es más relevante su procesamiento. Para ello se analizan y estructuran los datos y finalmente se muestran (Blanco, Teruel & Martín, 2016, p. 15).

Es importante mencionar que, como se aprecia a lo largo de esta investigación, el uso de Internet ha cambiado el ejercicio periodístico. Grandes medios como el diario británico *The Guardian* y *La Nación* de Argentina tienen equipos multidisciplinarios para el trabajo con datos, que incluyen a periodistas, desarrolladores web, ingenieros, diseñadores, etc. El equipo, además de ser multidisciplinario, también se caracteriza por ser

colaborativo y tecnológicamente versátil (Spagnuolo, 2019). El periodismo de datos es, además, una herramienta útil para evitar centrar la información periodística en las notas de prensa que envían las entidades públicas y privadas.

En tanto, para Fallas (2018), el periodismo de datos no es una moda, sino una necesidad y una oportunidad. Además, explica sus seis pilares:

1) Datos estructurados. Se refiere a que la información contenida en una base de datos tiene que estar correctamente estructurada y ser comprensible para que sea procesada por una computadora. Ello permitirá hacer el cruce de información o realizar cálculos.

2) Aplicación de matemática básica y estadística descriptiva. El periodista de datos debe tener conocimientos de matemática y estadística para analizar la información y no llegar a conclusiones equivocadas.

3) Representación visual de la información. Este pilar es importante para mostrar la información que, por la gran cantidad de datos que contiene, sería complicada o hasta imposible difundir solamente en texto. El lector no solo observa la información brindada, sino que puede navegar en las visualizaciones y obtener sus propias conclusiones.

4) Reporteo. Se hace referencia a la rigurosidad y precisión que se debe tener al momento de salir a la calle y ver el comportamiento de los datos en el mundo real.

5) Explicación de la metodología. Es importante ser transparentes con el lector y que este conozca cómo se desarrolló la investigación periodística: los pasos que se siguieron o los criterios estadísticos que se aplicaron; con la finalidad de que otras personas también puedan replicar la metodología en otros trabajos o, incluso, mejorarla.

6) Acceso a las bases de datos. Las bases de datos empleadas en las investigaciones deben poder ser descargadas en formatos *Excel*, *TXT (textfile)* o *CSV* con el fin de que los usuarios puedan usar esos datos para otro análisis o hacer una verificación.

También es relevante distinguir que Crucianelli (2013b) menciona cuatro productos del periodismo de datos: los artículos basados en datos, de corta extensión a partir de grandes volúmenes de información; las visualizaciones interactivas, que pueden ser independientes o estar incluidas en los artículos basados en datos; los conjuntos de datos abiertos o set de datos, que tienen como objetivo abrir datos públicos en formatos reutilizables; y las aplicaciones de noticias o *News App*, que usualmente son trabajadas entre periodistas y tecnólogos, como programadores o desarrolladores, debido a que se dispone de grandes cantidades de datos y resulta complicado hallar una noticia si es que no se elabora esta aplicación que permite agrupar y analizar variables.

Pese a la importancia del periodismo de datos, este no se desarrolla en todos los medios de comunicación, ya sea por la estructura de las redacciones que no incluye la disciplina o el tiempo que deben invertir los periodistas en su trabajo y que suele no resultar productivo para los medios con fines de lucro. La periodista Ryann Grochowski Jones, de *Propublica*, comenta que en la organización donde labora un proyecto periodístico puede demandar varios meses o hasta años de trabajo (Cortés, Luengo & Elías, 2018).

2.2.1.1. Antecedentes del periodismo de datos.

Dos disciplinas anteceden al periodismo de datos y sirven de base para su desarrollo: el periodismo de precisión y el periodismo de investigación asistido por computadoras. La primera se hace conocida en 1973 gracias al libro *El periodismo de precisión. La introducción de un reportero a los métodos de las ciencias sociales*, del profesor Philip Meyer, pero es más notable aún en los años 80 del siglo XX al formar parte de las redacciones de grandes periódicos, principalmente en Estados Unidos, como *The Washington Post*, *Los Angeles Times*, *Usa Today* y *The New York Times*. Su desarrollo

supone un cambio en la metodología del periodismo porque se emplean técnicas de investigación usadas principalmente en la estadística y las ciencias sociales para recolectar datos que posteriormente con el apoyo de herramientas informáticas se analizan. De acuerdo a Meyer (1993), los datos en bruto no son suficientes; por lo tanto, “para que sean útiles e inteligibles han de ser procesados, conceptualizados, integrados en algún esquema. Hay que insertar el dato material en algún marco de referencia mental que ayude a su interpretación y comprensibilidad” (p. 33-34). En el periodismo de precisión, entonces, se hace referencia a un periodismo que parte de la investigación tradicional, pero con un entorno más informático y en el que se usan nuevas tecnologías para publicar un producto donde no solamente se busca la inmediatez y se limita a brindar cifras.

La segunda disciplina es el periodismo de investigación asistido por computadoras, que puede ser catalogado como la evolución del periodismo de precisión en donde se incorporan al proceso investigativo aparatos tecnológicos como la computadora. De acuerdo a López (2016) también se emplea la hoja de cálculo como parte de la fase para procesar información; y, por supuesto, sin dejar de lado, según Segnini (2008), las técnicas del reporte tradicional.

López (2019) expone varios ejemplos de investigaciones que emplearon el periodismo asistido por computadoras. Por ejemplo, el reportaje titulado *Una mirada científica al crimen de Dade* (Florida, Estados Unidos), donde se analizaron tres mil sentencias criminales, con la participación de estudiantes de derecho y un programador, y se evidenció que la mayoría de detenidos por crímenes graves no fue encarcelado. Como este, hay decenas de casos que explican el trabajo periodístico con computadoras y análisis de datos. López (2019) también narra el caso de Elliot Jaspin, un periodista del *Providence Journal*, quien en 1987 descubrió que varios conductores de buses escolares en *Rhode Island* (Estados Unidos) tenían antecedentes por narcotráfico y sanciones de tránsito, tras

cruzar cinco mil registros de choferes con las bases de datos de procesos penales y accidentes de tránsito.

Existen varias ventajas del periodismo asistido por computadora. Por ejemplo, Segnini (2008) asegura que “irradia particularmente el ejercicio del periodismo investigativo porque multiplica las posibilidades para generar ideas para proyectos de investigación, minimiza la dependencia en las fuentes tradicionales y permite al reportero obtener y entrevistar universos de datos para descubrir hechos ocultos” (p. 46).

Además, identifica cuatro funciones del periodismo asistido por computadoras: “generación de ideas, referencia, reporte y análisis y presentación” (Segnini, 2008, p. 47). En relación a la primera función, asegura que para poder detectar temas prometedores es importante mantener conversaciones periódicas con fuentes clave. De ahí, que recomiende los grupos de noticias, foros, listas de correo, blogs, agregadores *RSS (Really Simple Syndication)* y chats para mantener una comunicación activa. La segunda función del periodismo asistido por computadora es que permite hallar información de referencia en páginas web, discos, enciclopedias, entre otros productos elaborados por personas, que ayuden a complementar la investigación que realiza un reportero. Es decir, es el material que brinda datos o definiciones útiles para la actividad periodística.

La tercera función que se menciona es la del reporte donde el periodista aplica las técnicas principales del periodismo para la realización de su labor y que incluye, por ejemplo, la contrastación de fuentes o entrevistas. Para este proceso, “los periodistas pueden valerse de hojas electrónicas para analizar largas series de registros y realizar cálculos” (Segnini, 2008, p. 52), y también emplear programas estadísticos. Es importante recalcar que cuando se aplican las técnicas del periodismo asistido por computadoras el periodista está en la capacidad de acceder a fuentes de primera mano, correspondencia oficial, información en diversos soportes como documentos, fotografías, videos o audios.

Finalmente, en la función de análisis y presentación es importante que el periodista recurra a programas o herramientas que le permitan encontrar patrones y llegar a conclusiones puntuales. En estos casos, se menciona el uso de los mapas conceptuales, el análisis de redes sociales y la georreferenciación.

2.2.1.2. Ámbitos del periodismo de datos.

Para analizar cómo se integra el periodismo de datos en los medios de comunicación, de acuerdo a la explicación de Bueno (2014), es necesario abordar cuatro esferas del periodismo: “el ámbito empresarial, profesional, tecnológico y de contenido” (p. 9).

En el ámbito empresarial se hace referencia a los equipos de trabajo que deben ser multidisciplinarios para así completar las diferentes fases del proceso investigativo. Por ejemplo, se necesitan personas con habilidades en la investigación tradicional, con conocimientos en estadísticas, diseñadores y programadores.

El ámbito profesional, en tanto, involucra al perfil del periodista de datos. El hecho de aplicar diferentes metodologías de trabajo al periodismo de investigación, el manejo de abundante información y el uso de la tecnología ha cambiado el perfil del periodista. De acuerdo a Esteban (2012, p. 20) “cabén muchas descripciones en este perfil, pero se trataría principalmente de una persona capaz de tratar y analizar grandes bases de datos y, sobre todo, ser capaz de contar historias dentro de ellas”.

Entre las características que debe tener un periodista de datos, se encuentra el dominio del inglés ya que la mayoría de las herramientas que son útiles para el periodismo se encuentran en ese idioma, saber hacer búsquedas en la web profunda, tener conocimiento de *Excel* y de matemática (Crucianelli, 2013, pp. 123-124). Pero el perfil en realidad es mucho más complejo. Un periodista de datos tiene que conocer las

herramientas que le ayudarán a visualizar datos en caso no domine algún lenguaje de programación, debe tener mucha paciencia, ser minucioso para evitar los errores y detectar la noticia. Asimismo, saber interpretar los datos, conocer herramientas de limpieza y análisis de datos como *Open Refine*, conocido también como *Google Refine*, ser autodidacta porque son pocos los países en donde hay formación académica y tener conocimientos sobre transparencia y acceso a la información pública.

Finalmente, los ámbitos tanto tecnológico como de contenido se caracterizan por el uso de programas o herramientas informáticas en las fases del proceso de producción de la noticia y la manera en la que se va a presentar.

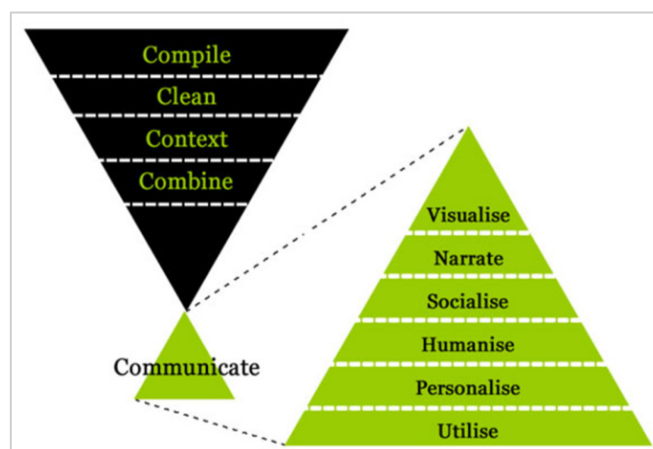
2.2.1.3. Metodología del periodismo de datos.

Bradshaw (2011) explica el proceso del periodismo de datos con un gráfico de una pirámide invertida: desde que se obtiene gran cantidad de información hasta el momento de comunicar los resultados. De acuerdo a Flores y Salinas (2013), el esquema muestra cómo se avanza en el trabajo con grandes cantidades de datos que finalmente serán presentados de la manera más clara a la audiencia.

En esta pirámide, que tiene como referencia bases ya explicadas por el profesor Philip Meyer, se detallan cinco fases para la generación de contenido: compilar (*compile*), limpiar (*clean*), contextualizar (*context*), combinar (*combine*) y comunicar (*communicate*), tal como se aprecia en la Figura 4. En la última fase, además, se explica la variedad de opciones que existen para difundir los datos: visualizar (*visualise*), narrar (*narrate*), socializar (*socialise*), humanizar (*humanise*), personalizar (*personalise*) y utilizar (*utilise*).

Figura 4

Pirámide invertida del periodismo de datos



Nota. Adaptado de *The inverted pyramid of data journalism*, por P. Bradshaw, 2011, *Online Journalism Blog* (<https://onlinejournalismblog.com/2011/07/07/the-inverted-pyramid-of-data-journalism/>).

Segnini (2012, como se citó en Ramírez, 2012) menciona cinco pasos para trabajar, de acuerdo a su presentación en la Conferencia Latinoamericana de Periodismo de Investigación (Colpin) en su edición de 2012¹: la obtención de datos, limpieza, análisis, verificación y visualización. A diferencia de Bradshaw (2011), ella incluye el análisis y la verificación de la información. Estas dos fases son importantes puesto que se debe interpretar la información recopilada para obtener conclusiones y corroborar que el contenido sea veraz. También recalca que el periodismo de datos no reemplaza la manera tradicional de ejercer la profesión, sino que hacer mejor el trabajo.

Es así que, considerando los aportes de Bradshaw (2011) y Segnini (2012, como se citó en Ramírez, 2012), se establece para esta investigación que la metodología de trabajo en el periodismo de datos está compuesta por siete procesos que permiten su mejor desarrollo y aplicación, y que se detallan a continuación.

¹ <http://www.puoperiodismo.cl/revisa-el-programa-de-la-conferencia-latinoamericana-de-periodismo-de-investigacion-2012/>

1. Compilar. La compilación es el proceso de recolección de información y puede ser de varias formas: información entregada directamente por una institución, encontrada por búsquedas avanzadas en Internet; mediante un *script*, que de acuerdo a Hidalgo y Torres (2016) es “un programa simple que permite bajar información de manera automatizada” (p. 19) de sitios web; recolectada por observación o encuestas, etc.; y en varios tipos archivos como *CSV*, *JSON (JavaScript Object Notation)*; *Excel*; *PDF (Portable Document Format)*; *PNG (Portable Network Graphics)*; *Word*, entre otros.

En este proceso de compilación los datos pueden ser obtenidos de fuentes privadas o públicas a través de bases de datos abiertos o conseguidas por medio de leyes de acceso a la información. Aunque para muchas personas puede resultar fácil, la recopilación de información en realidad no lo es porque existen funcionarios que no comprenden la intención de los periodistas de acceder a bases de datos y cómo les puede servir en su labor. De ahí que sea importante, según Flores y Salinas (2013), “insistir con todos los protocolos que permiten el acceso a la información” (p. 18).

2. Limpiar. La limpieza, según Bradshaw (2011), consiste en eliminar el error humano y convertir la información en una estructura útil para luego analizarla y cruzarla con otras bases de datos si es necesario. Esto implica corregir información mal escrita, formatos, y estructurar las bases de datos de una manera que puedan ser usadas por el periodista. El procedimiento de limpieza se puede realizar por medio de diferentes herramientas como las hojas de cálculo de *Microsoft Excel* o de *Google Drive*; *MongoDB*, un sistema de base de datos *NoSQL* (no solo *SQL*) para documentos de código abierto; con el lenguaje de programación *C++*; u *Open Refine*. Los archivos que se pueden importar para trabajar en esta última herramienta mencionada pueden tener las extensiones *TSV (Tab-Separated Values)*, *CSV*, *XML (Extensible Markup Language)*, *JSON* e incluso *Google Spreadsheets*, entre otros. También permite transformar archivos de cualquiera de

estos formatos a otro. *Open Refine* funciona como ejecutable sobre cualquier navegador web y está disponible para Windows, Mac y Linux (Ríos, 2014).

Además, los periodistas también utilizan con frecuencia *Tabula* y *OCR (Optical Character Recognition)*. La primera herramienta es de código abierto y permite a los usuarios subir archivos en *PDF* y extraer datos tabulares en *CSV*, a través de un interfaz amigable que incluso han usado los periodistas de *Propublica* en el proyecto *Dollar for Docs*² (Weiss, 2013). La segunda herramienta, el reconocimiento óptico de caracteres es un sistema que permite convertir *PDF* e imágenes en archivos *JPEG (Joint Photographic)*, *BMP (Bit Mapped Picture)*, *TIFF (Tagged Image File Format)* o *GIF (Graphics Interchange Format)* en formatos editables como *Word*, *Excel* y texto. *OCR* tiene versiones de paga y gratuitas en Internet sin la necesidad de instalarlo en la computadora. Ambas herramientas simplifican la labor periodística cuando la información se encuentra en formatos poco amigables con el fin de extraer datos y evitar el tipeo manual.

3. Contextualizar. En la parte de la contextualización es importante conocer la metodología con la que se obtuvieron los datos. En palabras de Fallas (2014), es necesario entrevistar a las bases de datos, ya que son una fuente más que es propensa a contar mentiras o brindar datos errados, de ahí que se deba verificar su autenticidad. En esta línea, recomienda hacer cuestionamientos como ¿quién hizo la recolección de información y con qué fin?, ¿cuál fue la metodología de recolección?, ¿qué tan confiable es la organización o persona que construyó la base de datos?, ¿la información está completa?, ¿con qué objetivos se elaboró la información? y ¿la información es suficiente o se requiere complementarla o validarla con otras bases de datos?

² *Dollar for Docs* es una investigación periodística que reveló los pagos que las empresas farmacéuticas y de dispositivos médicos realizan a médicos y hospitales de Estados Unidos por charlas de promoción, consultorías, investigaciones, entre otros aspectos.

También es relevante comprender todos los términos que aparecen en las bases de datos ya que, usualmente, algunos significados están representados por códigos o son muy técnicos y poco comprensibles para personas no especializadas en determinado campo de estudio. Si la información no cuenta con un diccionario de datos que explique su contenido, el periodista no debe dudar en solicitar una entrevista con la persona que hizo la recolección de la información o que maneja y entiende esos datos con el fin de despejar interrogantes. Asimismo, debe considerar no solo centrarse en el dato bruto, sino en qué otros factores son importantes para su posterior análisis.

4. Combinar. Una noticia con más de una fuente siempre va a ser más consistente, sobre todo porque ayuda a encontrar relaciones que una sola base de datos no podría otorgar a primera vista. Por ejemplo, si se tiene un registro de políticos acusados de corrupción con sus números de documento nacional de identidad (DNI) y otra base de datos de propiedades, también con números de DNI o nombres, se podría saber a qué políticos le corresponde cada propiedad. Este caso ejemplifica el cruce de bases de datos y los hallazgos dependen “de la calidad de preguntas y de la combinación de dos o más registros para encontrar coincidencias reveladoras” (Hidalgo y Torres, 2016, p. 20).

5. Analizar. En esta fase de trabajo, el periodista compara la información que posee, cruza datos y busca patrones de comportamientos que no suelen ser típicos y hasta encuentra tendencias. Aquí, las herramientas que se suelen emplear incluyen *Excel* y *Open Refine*. También, los lenguajes de programación *R*, *Python* y *SQL (Structured Query Language)*, así como *Neo4j*, una herramienta que permite identificar conexiones en grandes bases de datos y verlas en gráficos. Estos últimos lenguajes y herramientas son,

principalmente, usados por los tecnólogos. Justamente, *Neo4j* fue empleada en la serie periodística *Swiss Leaks*³.

6. Verificar. Según explican Hidalgo y Torres (2016), en esta fase de trabajo el periodista aplica la metodología tradicional del periodismo de investigación y entrevista a las personas involucradas, busca nuevos documentos que contrasten las bases de datos para hallar debilidades o confirmar sus fortalezas. Por ejemplo, cuando las instituciones públicas brindan documentos u hojas de cálculo es posible que haya errores de tipeo o conclusiones, puesto que la información ha sido elaborada por humanos.

7. Comunicar. Tiene como objetivo mostrar la información y hallazgos. Ahora con el avance de la tecnología hay varias formas de presentar la noticia: se pueden usar visualizaciones interactivas, gráficos, videos, audios, etc. El uso del formato dependerá mucho del tipo de información que se tenga y el tipo de historia que se quiera contar. A continuación, se enumeran las seis maneras de comunicar, según Bradshaw (2011):

- Visualizar. Es uno de los caminos más rápidos de mostrar los resultados de las investigaciones periodísticas, a pesar de que no resulta un proceso sencillo. El periodista debe encontrar la manera de explicar de la mejor manera posible uno o varios datos. Para ello, existen decenas de herramientas, en versiones gratuitas y de paga, que los reporteros pueden utilizar. Hidalgo y Torres (2016, p. 51) refieren seis: *Datawrapper*; herramienta con la que se pueden convertir datos de una hoja de cálculo en gráficos o mapas; *Infogram*, que permite elaborar infografías luego de insertar los datos o importarlos en archivos como *Excel* o *CSV*; *StoryMapJS*, que crea historias con la identificación de lugares específicos y la incorporación de imágenes, videos o texto; *TimelineJS*, que permite la elaboración de

³ Proyecto colaborativo internacional impulsado por el Consorcio Internacional de Periodistas de Investigación que contó con la participación de medios como *CBS*, *Le Monde*, *The Guardian*, *The Washington Post* y *Süddeutsche Zeitung*. La investigación reveló que la filial suiza del Banco HSBC creó un sistema de evasión fiscal para que más de 100 mil clientes con gran adquisición económica, en 203 países, escondan su dinero y así eviten el pago de impuestos entre el 2005 y 2007. El 2015 ganó los *Data Journalism Awards*.

líneas de tiempo con imágenes, mapas, hipervínculos o videos; *Tableau Public*, que también convierte los datos de las hojas de cálculo en gráficos interactivos; y *Soundcite*, útil para insertar sonido a una palabra o frase.

Es importante destacar que, para el uso de cualquiera de estas herramientas, intuitivas por naturaleza, no es indispensable conocimientos de programación. Además, así como existen herramientas utilizadas para la creación de visualización interactivas, también hay aplicaciones informáticas para crear gráficos estáticos como *Illustrator* y *Photoshop*, empleados por los periodistas, diseñadores o programadores.

- Narrar. No se trata solamente de mostrar los datos, sino de encontrar la manera de plasmarlos para que puedan ser entendidos por los lectores, como las comparaciones. Por ejemplo, si se desea mencionar un monto alto de dinero podría hacerse el cálculo en sueldos mínimos vitales o si se habla de una superficie, hacer la relación con lugares conocidos: ciudades, parques o estadios.

- Socializar. La recolección de información a través del contacto con la comunidad también puede brindar una dimensión social a los datos.

- Humanizar. Se trata de mostrar los datos por medio de historias reales. Buscar un caso o una persona que ejemplifique la información con el fin de que los lectores no solo lean números, y se sientan identificados con lo que se revela.

- Personalizar. Esta manera de comunicar involucra al lector con situaciones de su vida. Por ejemplo, se puede crear una aplicación sobre presupuestos públicos para que cada usuario acceda a la información de acuerdo a los datos personales que ingrese, como su lugar de residencia. Así hay un mayor involucramiento entre el lector y la investigación.

- Utilizar. Se refiere a comunicar los resultados con una herramienta que almacene la información empleada, como las calculadoras o las herramientas con Sistema de Posicionamiento Global (*Global Positioning System - GPS*).

2.2.2. La Ley de Transparencia y Acceso a la información Pública

La Ley fue publicada el 3 de agosto de 2002 durante el gobierno del expresidente Alejandro Toledo Manrique, pero entró en vigencia en enero de 2003. Tiene como finalidad “promover la transparencia de los actos del Estado y regular el derecho fundamental del acceso a la información consagrado en el numeral 5 del artículo 2 de la Constitución Política del Perú” (D. S. N° 021-2019-JUS, 2019).

Todas las personas tienen el derecho de pedir y recibir información por parte de las entidades públicas sin la necesidad de que el funcionario encargado de aceptar el documento le pregunte al solicitante por qué está haciendo el pedido ni mucho menos qué hará con los datos obtenidos. Sin embargo, estas preguntas suelen ser recurrentes en las mesas de partes presenciales al momento de dejar pedidos.

Se entiende como información pública todo documento “financiado por el presupuesto público que sirva de base a una decisión de naturaleza administrativa, así como las actas de reuniones oficiales” (D. S. N° 021-2019-JUS, 2019). De acuerdo al Reglamento de la Ley⁴, cualquier persona natural o jurídica puede presentar el pedido ya sea a través de los formularios virtuales implementados por las instituciones, en su mesa de partes o a las direcciones de correo electrónico que consignen. Sin embargo, debe considerar colocar ciertos datos al momento de llenar su solicitud: nombres y apellidos, número de identificación en caso de ser mayores de edad, número de teléfono, correo electrónico, firma o huella digital, “expresión concreta y precisa del pedido de

⁴ Aprobado mediante Decreto Supremo N° 072-2003-PCM

información” o cualquier otro dato que ayude a la su búsqueda; modalidad en la que se desean recibir los datos; y dependencia que posee la información (D. S. N° 072-2003-PCM, 2003). La Figura 5 es un modelo referencial para solicitar información pública en caso no se opte por presentar los formularios de cada institución pública.

Figura 5

Modelo de solicitud de acceso a la información pública

SOLICITUD DE ACCESO A LA INFORMACIÓN PÚBLICA	
Nombre de la Institución	
I. Funcionario encargado de entregar la información:	
II. Datos del solicitante:	
Nombres y apellidos o razón social:	Documento de identidad y número:
Dirección:	
Correo electrónico:	Número telefónico:
III. Información solicitada:	
IV. Dependencia a la que se solicita la información:	
V. Forma de entrega de la información	
Nombres y apellidos:	Fecha:
Firma:	
Observaciones: _____	

Aunque la ley no determina un formato oficial de solicitud, este modelo ayuda a pedir información pública.

Nota. Adaptado de *La navaja suiza del reportero* (p. 93), por D. Hidalgo y F. Torres, 2016, *Asociación de Periodismo de Investigación Ojo Público y Consejo de la Prensa Peruana*.

Todos los pedidos deben ser dirigidos a los funcionarios que las entidades estipulan como responsables de la entrega de información. Esto usualmente se aprecia en los portales virtuales de transparencia. En caso de que no hubieran sido designados, la solicitud se envía al funcionario que tiene en su poder los datos o a su superior inmediato. Si no se

tuviera ese dato o se hubiera completado erróneamente, la institución debe derivar el pedido a la persona competente para atenderlo, incluso si el solicitante no especificó el nombre de la dependencia que posee la información.

Es importante mencionar que las entidades públicas deben entregar la información contenida en diversos formatos como fotos, documentos escritos, en soporte digital, entre otros, siempre que haya sido creada u obtenida por la institución o esta la tenga en su poder. Y solo deben cobrar por el soporte en el que se entrega la información, como la copia o los discos.

Cuando la Ley entró en vigencia, el plazo máximo para que las instituciones respondieran las solicitudes era de siete días hábiles, pero tras una modificatoria el 2019 se estipuló que sean diez. Aquí hay varias consideraciones que se deben tener en cuenta. Por ejemplo, si la entidad pública no puede cumplir con este plazo de entrega por falta de capacidad logística, de personal o porque la información es abundante, puede establecer una nueva fecha de entrega, sin límite específico, siempre y cuando lo informe dentro de los dos días hábiles de haber recibido el pedido de información.

Si los funcionarios incumplen las disposiciones de la Ley, podrían ser sancionados o denunciados penalmente por abuso de autoridad.

2.2.2.1. Denegatoria de información.

Las entidades públicas deben sustentar las razones por las que niegan información y nunca justificar este hecho basándose en la identidad del solicitante. De acuerdo al Texto Único Ordenado de la Ley, hay tres tipos de datos que no pueden ser revelados: los secretos, reservados y confidenciales.

La información secreta es aquella que muestra aspectos de seguridad nacional, o que está relacionada a garantizar la seguridad de las personas, y cuya difusión pone en

riesgo la democracia o la integridad del territorio. Por ejemplo, dentro de este grupo se encuentra la información clasificada en los ámbitos militar y de inteligencia. Según la norma, esta información podría ser entregada recién cinco años después de su clasificación, siempre que ello no implique riesgos.

El otro grupo corresponde a los datos reservados. Aquí se incluye la información relacionada a la lucha contra el crimen en el país, como los planes de operaciones policiales, el armamento, negociaciones internacionales o información que podría perjudicar las relaciones internacionales.

Finalmente, en el tercer grupo se ubica la información confidencial. Dentro de las excepciones destacan los datos protegidos por el secreto bancario, industrial, tecnológico, bursátil y tributario; información vinculada a investigaciones que se encuentren en curso; que podría revelar la estrategia que se sigue en un proceso judicial o administrativo; protegida por el secreto profesional; y la información que incluye datos personales y cuya difusión invade la intimidad personal y familiar. Los datos personales son “toda información sobre una persona natural que la identifica o la hace identificable a través de medios que pueden ser razonablemente utilizados” (Ley N° 29733, 2011, p. 1).

En caso se solicite documentación que contiene de forma parcial información relacionada a las excepciones mencionadas anteriormente, las entidades de la administración pública deben entregar los datos accesibles del documento. Sin embargo, la práctica muestra que las instituciones públicas no suelen hacer ello y optan por negar toda la información, así como no respetar los plazos para entregar la información solicitada, entre otros aspectos relacionados al correcto uso de la Ley (Ministerio de Justicia y Derechos Humanos, 2022).

También es importante recordar que las instituciones del Estado no están obligadas a crear información que al momento de la solicitud no disponen ni mucho menos interpretar o hacer análisis de su documentación para emitir una respuesta.

Cuando el pedido es negado, el organismo del Estado deberá explicar las razones por las cuales no entrega la información, la que debe estar amparada en la Ley. Si esto no ocurre, el solicitante tiene el derecho de apelar la decisión de la institución. Asimismo, si la persona que pidió información no recibe respuesta dentro de los diez días hábiles, puede interpretar el pedido como denegado y presentar un recurso de apelación por la vía administrativa o recurrir a la vía judicial mediante un hábeas data. Esto también aplica en caso de que los datos sean ambiguos o no cumplan las expectativas del solicitante.

2.2.2.2. El recurso de apelación.

Este documento deberá incluir un detalle del proceso de solicitud de información: qué se pidió, en qué fecha, qué tipo de datos se obtuvieron, la respuesta recibida y en caso de haber sido negada, explicar cuál fue la justificación recibida. Además, se puede detallar de manera opcional por qué razones se considera que la respuesta debe ser brindada, como lo realiza el Centro Liber⁵, un proyecto del Instituto Prensa y Sociedad, dedicado al tema de transparencia y acceso a la información pública.

El documento puede ser entregado a la entidad pública a la que se solicitó la información dentro de los quince días calendarios de haber recibido una respuesta negativa para que la remita al Tribunal de Transparencia y Acceso a la Información Pública (TTAIP), un órgano dependiente del Ministerio de Justicia y Derechos Humanos. En caso la institución pública no haya brindado respuesta, el solicitante de información no tiene un plazo límite para presentar este recurso. Cabe mencionar que la apelación también puede

⁵ https://www.centroliber.pe/es/transparencia?title=&filtro_categoria=2

ser enviada directamente al TTAIP, organismo que en cualquiera de los casos tiene, desde que admitió el recurso, diez días hábiles para emitir una respuesta. Sin embargo, este plazo no siempre se cumple debido a la alta carga procesal que tiene.

El TTAIP es la última instancia administrativa para apelar. Si la resolución fue desfavorable para el solicitante de información, este puede recurrir a un hábeas data como última opción. También puede hacerlo directamente sin pasar por la vía administrativa, pero debe tener en cuenta que es un proceso de varios años que incluso puede terminar de resolverse en el Tribunal Constitucional.

2.2.2.3. El hábeas data.

El hábeas data⁶ “es una garantía que protege dos derechos fundamentales: el derecho a la información y la autodeterminación informativa o protección de datos personales” (Quiroz, 2016, p. 24). En este contexto, según Perla (2010), hasta hace pocos años se demostraba que el hábeas data había sido usado, principalmente, para acceder a información pública en vez de la protección de datos personales. Para poder presentar esta acción judicial es necesario que el ciudadano haya realizado una solicitud de acceso a la información pública y esta no haya sido respondida en los plazos estipulados por Ley o, en todo caso, la información esté incompleta o sea ambigua.

Para seguir con este procedimiento, tal como detallan Hidalgo y Torres (2016), el ciudadano no está obligado a ser representado por un abogado para dejar la documentación, pero sí tendrá sesenta días hábiles, luego de que le notificaran la negativa o venciera el plazo de respuesta, “para presentar un hábeas data ante el juez civil o el juez mixto de la zona donde vive o del lugar donde se ubica la institución pública que negó su pedido” (p. 92).

⁶ Este recurso se incorporó en la Constitución Política del Perú de 1993.

La demanda, en el mejor de los casos, se resuelve en seis meses. En la peor de las situaciones puede demorar varios años, como en el caso del hábeas data que siguió *Ojo Público*, en el contexto del proyecto *Memoria robada*, desde el 19 de agosto de 2016 contra del Ministerio de Cultura por los certificados de exportación de réplicas de bienes inmuebles pertenecientes al patrimonio nacional. El 2021 recién se emitió un fallo⁷ desde el Tribunal Constitucional a favor del medio nativo digital.

2.2.2.4. El periodismo de datos y el acceso a la información pública.

En el periodismo de datos, como ya se ha detallado anteriormente, se manejan grandes cantidades de información que son recopiladas, limpiadas, contextualizadas, combinadas, verificadas, interpretadas y comunicadas. Estos datos, en gran parte de los casos, pueden ser obtenidos a través de la Ley, un mecanismo que hubiera sido inimaginable de utilizar hace veinte años en el país.

De acuerdo a Magallón (2013, p. 6), “el derecho de acceso a la información permite que surjan disciplinas como el periodismo de datos”. Y con la información recopilada y posteriormente publicada incrementan las posibilidades de fiscalizar a las entidades públicas. Cuando un periodista solicita información existe la posibilidad de que esta llegue a más personas ya sea porque la publica, menciona o crea herramientas didácticas para que la ciudadanía se informe sobre un determinado tema. De aquí nace la importancia de que se promueva el acceso a la información pública y también de que las entidades del Estado empleen mecanismos para crear y almacenar la información con la que laboran.

La Ley puede ser catalogada como una herramienta de trabajo para el periodista de datos. Pero para ello debe estar capacitado o tener nociones básicas sobre los siguientes

⁷ Expediente N° 02422-2019-PHD/TC

aspectos: conocer la normativa, saber cómo apelar una respuesta disconforme, estar informado sobre los datos que maneja cada entidad pública o en su defecto tener disposición para consultar previamente sobre ellos y promover el uso de la Ley tanto en el campo periodístico como en el ciudadano. A esto se le suma que también deberá contar con habilidades para procesar esa información, limpiarla, interpretarla y verificarla para poder comunicar hechos poco difundidos.

Cuando un periodista de datos hace uso de la normativa, y de alguna manera libera información a la ciudadanía, sirve como fiscalizador del Gobierno, promueve la participación pública, brinda un arma contra la corrupción, difunde conocimientos a los ciudadanos para que tomen mejores decisiones, y corrige errores si los hay en el contenido de la información que consigue. Si bien en Perú la Ley es útil para el periodismo de datos, en el transcurso de los años se han apreciado trabas para los periodistas, debido a la negativa de entregar información.

2.2.3. Los medios nativos digitales peruanos Ojo Público y Convoca

El periodismo de datos en el Perú tiene antecedentes en una investigación sobre corrupción que realizó a fines de 1997 Liz Mineo, entonces periodista del diario *El Comercio*. Tras el Fenómeno del Niño, ella cuestionó cómo el Gobierno gastó el presupuesto de S/ 100 millones para revertir las consecuencias causadas por el desastre. El dinero estaba administrado por el Instituto Nacional de Defensa Civil (Indeci), según explican Hidalgo y Torres (2016).

A falta de normas de acceso a la información pública, Liz Mineo convenció a un funcionario del Indeci para que le brinde información sobre las obras y contratistas involucrados en remediar los daños causados por el Fenómeno del Niño. Es así que recibió ochenta páginas en físico de información que luego tuvo que procesar a una hoja de

cálculo. De esta manera, la periodista detectó varias irregularidades, como que diferentes obras de prevención estaban ubicadas en regiones que no eran vulnerables frente a este fenómeno climático. También halló que doce empresas contratadas eran de propiedad de un amigo del entonces general del Indeci, Homero Nureña; y que otras compañías estaban a nombre del sobrino y secretaria de este funcionario.

Además de esta publicación, hay otras investigaciones que muestran indicios del periodismo de datos en el país, sobre todo a partir de la entrada en vigencia de la Ley de Transparencia y Acceso a la Información Pública. En 2010, la periodista de *IDL-Reporteros*, Milagros Salazar, hizo una investigación sobre la industria pesquera en base a 47 mil reportes de desembarque de anchoveta, y detectó un sistema fallido de auditoría que beneficiaba a la industria pesquera y que representaba un aproximado de US\$ 100 millones en impuestos no pagados.

Es a partir de 2014 cuando se puede hablar de periodismo de datos en sí con el surgimiento de los medios nativos digitales *Ojo Público* y *Convoca*. Ambos explican, abiertamente, que ejercen el periodismo de datos y que usan la tecnología para contar historias periodísticas diferentes a las que se muestran en los medios tradicionales y masivos de comunicaciones. Sus redacciones, además de contar con periodistas de investigación, están conformadas por periodistas de datos, programadores, fotógrafos y desarrolladores web. Desde su lanzamiento, los dos medios han publicado varias investigaciones, ya sea con su propio equipo o en alianzas con medios locales y extranjeros, y organizaciones no gubernamentales. Entre los proyectos publicados en los primeros años de la creación de ambos medios, en los que se han realizado mayores cantidades de solicitudes de acceso a la información pública y que han sido reconocidos internacionalmente, se encuentran *Memoria robada*, de *Ojo Público*, y *Excesos sin castigos*, de *Convoca*; ambas objeto de estudio en esta investigación.

Los periodistas Óscar Castilla, Nelly Luna, Fabiola Torres y David Hidalgo fundaron la *Asociación de Periodismo de Investigación Ojo Público* el 2014. Los tres primeros personajes en mención trabajaron en el diario *El Comercio* por cerca de diez años e integraron la Unidad de Investigación del medio. El 27 de marzo de 2014 se les comunicó que el área en la que laboraban iba a cerrar y que, por lo tanto, cada uno de ellos iba a ser designado a otros equipos. No obstante, ellos optaron por renunciar y emprender su propio medio de comunicación. A esta iniciativa se sumó el periodista independiente David Hidalgo.

Tras varios meses de coordinaciones, luego de tener acercamiento con las comunidades de programadores y de haber realizado una alianza con la organización *Suma Ciudadana*, los periodistas lanzaron su primer proyecto de investigación: *Cuentas juradas*. Esta publicación mostró el análisis de las declaraciones juradas de los candidatos a la alcaldía de Lima que buscaban la reelección. El impacto de la investigación fue grande y empezó a republicarse en los medios de comunicación tradicionales y en diferentes portales web periodísticos. El 2015, *Ojo Público* recibió los *Data Journalism Awards* por *Cuentas juradas* como Mejor investigación del año en la categoría *Small Newsroom*.

A partir de esta investigación, el medio se hizo conocido y también comenzó a cubrir otras temáticas. Posteriormente lanzó la aplicación periodística *Cuidados intensivos*⁸, *Suprema fortuna*⁹, los *Panama Papers*¹⁰, investigación liderada por el Consorcio Internacional de Periodistas de Investigación y el diario alemán *Süddeutsche Zeitung*, *Fondos de papel*¹¹, *Implant Files*¹², entre otras.

⁸ Aplicación periodística que muestra el perfil detallado de los médicos en Perú y de las compañías privadas que prestan servicios de salud.

⁹ Investigación que revela la evolución patrimonial y perfil de carrera de más de 1,300 magistrados de la Corte Suprema y de los juzgados de Lima y Callao.

¹⁰ Investigación a partir de documentos filtrados que descubre nexos entre la firma panameña y Mossack Fonseca y personas o compañías en el Perú con el fin de crear sociedades offshore en paraísos fiscales.

¹¹ Plataforma que difunde información sobre los aportes que reciben y declaran los partidos u

El mismo año que se fundó *Ojo Público* también se constituyó la *Asociación Convoca*, un medio nativo digital que cuenta con investigaciones como la serie *Investiga Lavajato*, *Audios de la corrupción*, *El verificador de promesas*, etc. Según se observa en la página web de este medio, *Convoca* fue fundado por analistas de datos, reporteros y programadores bajo el liderazgo de la periodista Milagros Salazar, quien había laborado anteriormente en el diario *La República* y en el portal de investigación *IDL-Reporteros*. La primera publicación que elaboró *Convoca* fue *Swiss Leaks* en alianza con el Consorcio Internacional de Periodistas de Investigación.

Luego, desde el 2020, se han ido creando unidades de periodismo de datos dentro de los medios tradicionales, como *El Comercio Data* y *La República Data*.

2.2.3.1 El proyecto periodístico *Excesos sin castigo*.

*Excesos sin castigo*¹³ es el primer proyecto periodístico propio de *Convoca*; es decir, en el que no participó ningún otro medio de comunicación, y que inició en julio de 2014. La periodista Milagros Salazar¹⁴ detalla que la idea de elaborar esta investigación surgió luego de la aprobación de la Ley N° 30230, conocida como el paquetazo ambiental durante el gobierno del entonces presidente de la República, Ollanta Humala, ya que brindaba una especie de amnistía ambiental a las empresas con el objetivo de promover la inversión minera, de industrias extractivas y petróleo. Esto significa que, si había un proceso sancionador del Organismo de Evaluación y Fiscalización Ambiental (OEFA), una instancia dependiente del Ministerio del Ambiente, contra empresas de los sectores mencionados anteriormente, las multas por infringir las leyes ambientales quedaban

organizaciones políticas para sus campañas políticas con miras a las elecciones presidenciales.

¹² Serie periodística que analiza la industria de dispositivos médicos a nivel global y detecta prácticas abusivas de las compañías contra los pacientes, corrupción y conflictos de interés.

¹³ <https://excesosincastigo.convoca.pe/>

¹⁴ Comunicación personal, 7 de setiembre de 2019.

suspendidas cuando el proceso estaba en la primera instancia administrativa; mientras que, si se encontraba en segunda instancia, el pago se reducía en 50 %; salvo excepciones en donde esté en peligro la vida o la salud de las personas.

A raíz de este contexto, el quipo de *Convoca* se hizo varias interrogantes; por ejemplo, si la norma había sido una propuesta de las empresas y si es que había algún sector beneficiado principalmente; no obstante, al no hallar elementos suficientes para desarrollar esas hipótesis, optó por documentar el impacto de la norma y saber a cuánto ascendía el monto de las multas impagas. El monto resultante fue de S/ 30 millones.

El proyecto consiguió US\$ 50 mil de financiamiento de la organización *Oxfam* para desarrollar la investigación aproximadamente en ocho meses. El dinero se desembolsó a través de la empresa *IDL*, que cobró cerca del 15 % por la administración del fondo, con la finalidad de pagar los gastos operativos. Además, el financiamiento sirvió como un fondo semilla a favor de *Convoca* que le permitió contratar reporteros y programadores en el medio. En este contexto, resulta relevante mencionar que esta investigación periodística contó con la participación de diecinueve personas para su lanzamiento: cinco reporteros, tres programadores, tres periodistas asesores y ocho voluntarios en la sistematización de información. Sin embargo, solo los cinco reporteros y tres programadores integraron el equipo que le dedicaba tiempo completo a las actividades requeridas en el proyecto.

Como consecuencia de la investigación luego de dos años de haber sido publicada, el Congreso de la República derogó el artículo de la norma que brindaba privilegios para el sector. *Excesos sin castigo* fue finalista en el Premio Gabo 2016. Ese mismo año ganó los *Data Journalism Awards* en la categoría Aplicaciones de Noticias.

2.2.3.2 El proyecto periodístico *Memoria robada*.

Memoria robada es una plataforma periodística de *Ojo Público* que reveló información sobre el tráfico ilícito de patrimonio cultural en países latinoamericanos. La investigación, en su lanzamiento, fue liderada por *Ojo Público* y también contó con la participación de periodistas y programadores de los medios *Plaza Pública* (Guatemala), *La Nación* (Costa Rica), *Animal Político* (México) y *Chequeado* (Argentina); ya que se necesitaba una mirada global para conocer las rutas de tráfico y el impacto que generaba en Latinoamérica.

Esta investigación transnacional tuvo como objetivo comprender el flujo de bienes culturales entre diversos países de América Latina, debido a que el tema no se abordaba a profundidad en los medios tradicionales de comunicación; es decir, solo había publicaciones de casos aislados sobre piezas culturales robadas o repatriadas. Es a partir de este contexto que el equipo, conformado por cinco periodistas (incluyendo a la autora de la presente investigación) y un programador, que laboraron a tiempo completo, tres profesionales audiovisuales, y tres voluntarios en la sistematización de datos, realizó decenas de peticiones de acceso a la información pública a las entidades competentes en Perú para desarrollar el tema. Con las respuestas, el equipo creó bases de datos y publicó el primer buscador¹⁵ en el país que muestra 39 mil piezas de arte robadas en 132 países y reportadas a la Interpol en cien años, casas de subasta en catorce países que vendieron patrimonio cultural exportado irregularmente de países de Latinoamérica, más de 268 mil bienes registrados como patrimonio cultural en Perú, y la lista de 7 mil 500 bienes robados que veintidós países devolvieron. Para la presente investigación, se analizó el trabajo realizado en *Memoria robada* desde Perú.

¹⁵ <https://memoriarobada.ojo-publico.com/buscador/?categoria=robos&q=&p=1>

Memoria robada obtuvo el tercer lugar del Premio Latinoamericano de Periodismo de Investigación 2016, de Transparencia Internacional y del Instituto de Prensa y Sociedad (IPYS). Según narra la periodista Fabiola Torres¹⁶, este proyecto tuvo el apoyo económico de unos US\$ 24 mil por parte de la *International Women's Media Foundation* y de *Found for Investigative Journalism*.

¹⁶ Comunicación personal, 6 de setiembre de 2019.

CAPÍTULO III: HIPÓTESIS Y VARIABLES

3.1. Hipótesis

La metodología del periodismo de datos se aplicó parcialmente en los proyectos *Memoria robada*, de *Ojo Público*, y *Excesos sin castigo*, de *Convoca*; es decir, los periodistas y tecnólogos no desarrollaron todas las fases del proceso metodológico que incluye la recopilación, limpieza, contextualización, combinación, análisis, verificación y comunicación de la información.

3.2. Variables

La variable general es la metodología del periodismo de datos; mientras que las variables intermedias (dimensiones) son las siete fases que forman parte de la metodología que siguen los periodistas y tecnólogos en el ejercicio del periodismo de datos según Bradshaw (2011) y Segnini (2012, como se citó en Ramírez, 2012) — y que se detallan a continuación:

1. Recopilación de datos. Permitió conocer cuáles fueron las fuentes de información a las que accedieron los periodistas para elaborar los proyectos *Memoria robada* y *Excesos sin castigo*. Si bien la presente investigación está centrada en los datos obtenidos a través de la Ley, la recopilación también brinda una mirada global de las demás fuentes. Asimismo, se obtuvo detalles de las respuestas que dieron las entidades públicas a las solicitudes formuladas y el tipo de recurso que se aplicó cuando las respuestas fueron negativas o parciales: la apelación administrativa o el hábeas data.

2. Limpieza de datos. Ayudó a saber cuáles fueron las herramientas tecnológicas que emplearon los periodistas y tecnólogos de las investigaciones realizadas por *Ojo Público* y *Convoca*, así como los lenguajes de programación, y cuál era su nivel de

conocimiento. Asimismo, si participaron en la definición de estructuras de las bases de datos y transcripción de información.

3. Contextualización de datos. Permitió conocer si los periodistas analizaron a profundidad la información que obtuvieron antes de limpiarla y procesarla. Los indicadores considerados fueron los siguientes: explicación sobre los términos técnicos, conocer quién recolectó la información en la entidad pública, para qué finalidad se realizó esa recolección y cómo fue usada la información que se entregó. Esta fase de la metodología del periodismo de datos es fundamental para no hacer interpretaciones erradas, suposiciones ni mucho menos malos cálculos.

4. Combinación de datos. Tuvo como finalidad comprender la cantidad de bases de datos que se combinaron para hallar mejores datos e historias.

5. Análisis de datos. La aplicación de esta variable fue necesaria para conocer las herramientas tecnológicas y lenguajes de programación que utilizaron los periodistas y tecnólogos para interpretar la información que se recolectó, limpió, contextualizó y combinó.

6. Verificación de datos. Permitió determinar si se realizó un proceso de verificación de información y si fue de manera total o parcial.

7. Comunicación. Contribuyó a saber cuáles fueron los productos periodísticos que se publicaron; es decir, si fueron reportajes, aplicaciones periodísticas, visualizaciones, videos, audios, etc. También, las herramientas y programas empleados por los periodistas para visualizar datos; las herramientas y lenguajes de programación empleados por los tecnólogos para la creación del sitio web donde se alojaron las investigaciones de *Ojo Público* y *Convoca*, así como para visualizar información. Finalmente, esta variable permitió comprender las estrategias *online* y *offline* de difusión que se realizaron.

3.3. Operacionalización de Variables

Variable general	Variable intermedia (Dimensiones)	Categoría	Indicadores	Técnica de investigación
Metodología del periodismo de datos	Recopilación de datos	Vía de recolección de información	Aplicación de la Ley de Transparencia y Acceso a la Información Pública	Encuesta y entrevista
			Búsqueda avanzada en sitios web	Encuesta
			Entrevistas	
			Scraping	
		Recursos presentados para acceder a datos	Apelación administrativa	Encuesta
			Hábeas data	Encuesta y entrevista
	Limpieza de datos	Herramientas y lenguajes de programación	Excel	Encuesta
			Open Refine (Google Refine)	
			Tabula	
			OCR	
			R	
		Procesos	Definición de estructuras de bases de datos	Encuesta y entrevista
	Transcripción de información			
	Contextualización de datos	Información de las bases de datos	Responsable de recolectar la información	Encuesta
			Fecha de recolección	
			Razones de la recolección	
			Metodología de construcción	
			Significado de cada término	
			Diccionario de datos	
	Combinación de datos	Bases de datos	Cantidad de combinaciones	Encuesta
	Análisis de datos	Herramientas y lenguajes de programación	Excel	Encuesta
			Open Refine (Google Refine)	
			R	
Verificación de datos	Proceso	Total	Encuesta y entrevista	
		Parcial		
Comunicación de datos	Productos periodísticos	Artículos basados en datos	Encuesta	
		Visualizaciones interactivas		
		Conjunto de datos abiertos		
		Aplicaciones de noticias		
		Infografías		
		Videos		
		Audios		
	Herramientas y programas empleados por los periodistas para visualizar datos	Infogram	Encuesta	
		Datawrapper		
		Google Fusion Tables		
		HTML		
	Herramientas y lenguajes de programación empleados por los tecnólogos para visualizar datos y crear el sitio web	HTML	Encuesta	
		Java		
		Python		
		Indesign		
		Photoshop		
Ilustrator				
Estrategia online	Post para redes sociales	Encuesta y entrevista		
	Gráficos para redes sociales			
	Newsletter			
	Notas de prensa			
Estrategia offline	Entrevistas	Encuesta y entrevista		
	Conferencias			
	Talleres			

CAPÍTULO IV: MATERIALES Y MÉTODOS

4.1. Diseño de la investigación

Esta investigación es de tipo exploratoria porque abarca un tema que no se ha profundizado demasiado en Perú, desde el punto de vista académico. También es descriptiva ya que evalúa la aplicación de la metodología del periodismo de datos que siguieron los profesionales de los medios nativos digitales *Ojo Público* y *Convoca* en el desarrollo de los proyectos *Memoria robada* y *Excesos sin castigo*, respectivamente. Ambas investigaciones fueron elaboradas, principalmente, con información obtenida a través de la Ley de Transparencia y Acceso a la Información Pública.

De esta manera, el presente estudio tiene carácter mixto, cualitativo y cuantitativo, debido a que se aplicaron encuestas y entrevistas como técnicas para la recolección de información. A través de la aplicación de estas técnicas se pudo conocer el trabajo metodológico que siguieron los periodistas y tecnólogos de ambos medios en la realización de las investigaciones periodísticas. El mayor énfasis se encuentra en el ámbito cualitativo que, según Hernández, Fernández y Baptista (2010), “utiliza la recolección de datos sin medición numérica para descubrir o afinar preguntas de investigación en el proceso de interpretación” (p. 27).

4.2. Población y muestra

Se consideró necesario aplicar una muestra no probabilística porque la presente investigación busca describir y explicar la aplicación de la metodología del periodismo de datos en proyectos que tienen como componente principal datos obtenidos por medio de la Ley arriba mencionada. En ese sentido, fue indispensable conocer y analizar el trabajo de los tecnólogos y periodistas, incluyendo a los editores y coordinadores, que participaron en la ejecución de *Memoria robada* y *Excesos sin castigo*, para cumplir con ese objetivo.

Desde el lanzamiento de los portales web de *Ojo Público* y *Convoca* hasta el 2018, año en que empezó la presente investigación, cada medio publicó once y diez especiales periodísticos, respectivamente. En el caso de *Ojo Público* se identificaron *Cuentas juradas*, *Cuidados intensivos*, *Suprema fortuna*, *Fondos de papel*, *Audios de la corrupción*, *Dueños de la noticia*, *Memoria robada*, *Proyecto memoria*, *Panama Papers*, *Big Pharma* y *Oro sucio*. Por el lado de *Convoca* se encontraron *Investiga Lavajato*, *Audios de la corrupción*, *Confesiones en Sao Paulo*, *El verificador de promesas*, *Ilusión fiscal*, *La caída de Kuczynski*, *Memorias de la corrupción*, *Tierra herida*, *Sancionados* y *Excesos sin castigo*. Sin embargo, se consideraron para este estudio *Memoria robada* y *Excesos sin castigo* porque fueron los primeros proyectos periodísticos en los que se realizaron más solicitudes de acceso a la información pública: 117¹⁷ y 130¹⁸, respectivamente. Y también porque fueron desarrollados el 2016, a pocos años del lanzamiento de *Ojo Público* y *Convoca*.

El año en que se publicaron ambos proyectos periodísticos trabajaron en los medios en mención dieciocho personas a jornada completa, entre periodistas y tecnólogos. De este total, para conocer el trabajo metodológico empleado, se consideraron a aquellos que participaron directamente en la elaboración de *Memoria robada*, de *Ojo Público*, y *Excesos sin castigo*, de *Convoca* (ver Tabla 1).

En las dos investigaciones periodísticas también participaron voluntarios en la sistematización de la información; no obstante, debido a que las coordinaciones y ejecuciones de las tareas fueron realizadas eminentemente por los periodistas y tecnólogos, no se los incluyó en la muestra que sirvió como fuente de información. También se excluyeron a los tres profesionales audiovisuales de *Ojo Público* que colaboraron en la investigación, pero no participaron en la mayor parte del proceso investigativo.

¹⁷ Comunicación personal con Fabiola Torres, 6 de setiembre de 2019.

¹⁸ Comunicación personal con Milagros Salazar, 7 de setiembre de 2019.

Tabla 1*Población y muestra*

Población	Dieciocho periodistas y tecnólogos que trabajaron en <i>Ojo Público</i> y <i>Convoca</i> en el periodo en el que se ejecutaron los proyectos periodísticos <i>Memoria Robada</i> y <i>Excesos sin castigo</i> .	
Muestra	La muestra está constituida por trece periodistas y tecnólogos que participaron directamente en la elaboración de los proyectos periodísticos <i>Memoria robada</i> y <i>Excesos sin castigo</i> , dos de las investigaciones iniciales de los medios nativos digitales <i>Ojo Público</i> y <i>Convoca</i> , respectivamente. Cabe recalcar que los proyectos periodísticos fueron realizados principalmente con información obtenida a través de la Ley de Transparencia y Acceso a la Información Pública, donde ambos medios efectuaron en total más de doscientas solicitudes para recopilar información.	
	Periodistas	Tecnólogos
	<i>Ojo Público</i>	
	Fabiola Torres López, coordinadora general y editora.	Jason Martínez Vera, desarrollador.
	David Hidalgo Vega, editor general.	
	Johana Valer López, periodista asistente de investigación.	
	José Luis Huacles Unocc, periodista de investigación.	
	<i>Convoca</i>	
	Milagros Salazar Herrera, coordinadora general y editora.	Víctor Anaya Díaz, desarrollador.
	Aramís Castro Ramos, periodista y coordinador de tecnología.	Melissa Chávez Saldaña, desarrollador.
	Esteban Valle Riestra, periodista.	Luis Herrera Portella, desarrollador.
	Gabriela Flores Chávez, periodista.	
	Melanie Betetta Cabrera, periodista.	

4.3. Técnicas de recolección de datos

Como se ha mencionado, la presente investigación muestra una metodología mixta; es decir, cualitativa y cuantitativa, que desarrolló dos técnicas para recolectar datos: la encuesta y la entrevista. A todos los participantes encuestados y entrevistados se les informó que los datos brindados solamente serán usados en la presente investigación.

4.3.1. Encuesta

Esta técnica tuvo como objetivo recoger la información de los periodistas y tecnólogos que participaron directamente y a tiempo completo en la ejecución de las investigaciones periodísticas *Memoria robada* y *Excesos sin castigo*, de *Ojo Público* y *Convoca*, respectivamente. Los datos obtenidos permitieron describir la aplicación de la metodología del periodismo de datos que considera las siguientes etapas: recopilación, limpieza, contextualización, combinación, análisis, verificación y comunicación de la información. A través de las encuestas también se comprendió los niveles de conocimientos que tenían los periodistas y tecnólogos en diversas herramientas, programas y lenguajes de comunicación que se utilizaron durante determinadas fases de trabajo; y cómo adquirieron dichos conocimientos.

El instrumento que se aplicó fue el cuestionario (ver anexos 1 y 2) y específicamente estuvo dirigido a los periodistas que cumplieron a la vez las funciones de coordinadores y editores de los proyectos como Milagros Salazar, de *Convoca*; y Fabiola Torres y David Hidalgo, de *Ojo Público*. Asimismo, a los periodistas a tiempo completo Esteban Valle Riestra, Aramis Castro, Melanie Betetta y Gabriela Flores, de *Convoca*; y Johana Valer y José Luis Huacles, de *Ojo Público*. También participaron cuatro tecnólogos: Víctor Anaya, Melissa Chávez y Luis Herrera, de *Convoca*, y Jason Martínez, de *Ojo Público*.

La aplicación de los cuestionarios se realizó a través del *software* Formularios de Google. Estos contenían treinta y una preguntas para los periodistas y diecinueve para los tecnólogos, con la opción de marcar respuestas múltiples.

4.3.2. Entrevista

La entrevista se aplicó como una técnica complementaria a las encuestas y permitió conocer a profundidad detalles acerca de indicadores como la aplicación de la Ley de Transparencia y Acceso a la Información Pública y el hábeas data, dentro de la fase de la recopilación de datos; la definición de las estructuras de las bases de datos y la transcripción de información, en la etapa de limpieza de datos; el proceso de verificación de datos; y las estrategias *online* y *offline* aplicadas como parte de la comunicación de la información. De acuerdo a Kvale (2008), por medio de las entrevistas se obtienen “descripciones de situaciones y acciones específicas, no opiniones generales” (p. 35), de ahí la importancia de su uso.

Esta técnica cualitativa se aplicó a los coordinadores y editores de *Memoria robada*: Fabiola Torres y David Hidalgo, y de *Excesos sin castigo*: Milagros Salazar, quienes consintieron que sus nombres y respuestas sean utilizadas en esta investigación. Todas las entrevistas se realizaron presencialmente en junio y septiembre de 2019, fueron grabadas en audio y luego transcritas las respuestas (ver anexos 3, 4 y 5).

CAPÍTULO V: RESULTADOS

5.1. Presentación y análisis de los resultados

En este capítulo se presentan los resultados de los trece cuestionarios aplicados a los nueve periodistas y cuatro tecnólogos de *Ojo Público* y *Convoca* que participaron en el desarrollo de los proyectos periodísticos *Memoria robada* y *Excesos sin castigo*, respectivamente. Asimismo, de las entrevistas realizadas a los tres periodistas coordinadores y editores de los proyectos en mención.

Para la clasificación, tabulación y conteo de los resultados se utilizó la herramienta Formulario de Google que arrojó en una matriz las respuestas de los periodistas y tecnólogos que resolvieron los cuestionarios, y que permitió la elaboración de las figuras que se presentan en las siguientes páginas. En el caso de las entrevistas, estas fueron grabadas en audio y luego transcritas para extraer de ellas la información que permitió complementar este capítulo, en relación a los indicadores presentados, y otros datos que brindaron detalles generales de los proyectos periodísticos realizados.

En los cuestionarios además de realizar preguntas sobre el trabajo metodológico, también se consideraron algunas preguntas para conocer el perfil profesional de los periodistas y tecnólogos que participaron en ambos proyectos periodísticos.

Respecto a la información recopilada de los nueve periodistas, en el momento del desarrollo de *Memoria robada* y *Excesos sin castigo*, tres eran estudiantes de las carreras de Periodismo y Comunicaciones; cinco eran bachilleres, de Periodismo, Comunicaciones y Ciencia Política y Gobierno; y solo una persona tenía licenciatura en Comunicación Social. De todo este universo, un reportero no tenía experiencia en periodismo de datos al momento de empezar a laborar en los proyectos en mención; tres tenían menos de un año de experiencia y cinco tenían de entre uno a siete años. También se conoció que el 56 % de los periodistas tenía entre 20 y 25 años de edad; mientras que el resto superaba los 30 años.

En relación al perfil de los tecnólogos, todos desempeñaron el cargo de desarrolladores. El 75 % tenía el grado de bachiller en las carreras de Ingeniería de Sistemas y Ciencias de la Computación; mientras que el 25 % aún era estudiante de Ingeniería de Sistemas. De estos cuatro profesionales, solo uno tenía más de ocho años de experiencia en programación y el resto tenía desde meses hasta tres años como máximo. Además, el 75 % de los tecnólogos no tenía experiencia previa en periodismo de datos. Por último, las edades de estos cuatro profesionales oscilaban entre los 22 y 35 años.

Los resultados fueron agrupados en seis secciones que muestran el trabajo metodológico dentro del periodismo de datos: la recopilación de datos, la limpieza de datos, la contextualización y combinación de datos, el análisis de datos, la verificación de datos y la comunicación de datos. En algunos de los apartados, además, ha sido relevante la presentación de figuras para mostrar los resultados obtenidos que se presentan a continuación según cada fase del proceso metodológico.

5.1.1. Recopilación de datos

Para desarrollar las investigaciones de *Ojo Público (Memoria robada)* y de *Convoca (Excesos sin castigo)*, los equipos periodísticos obtuvieron información a través de entrevistas, búsqueda avanzada en sitios web y aplicando *scraping*, una técnica empleada para extraer información de sitios web con el uso de programas de *software*. El *scraping*, en ambos casos, no fue ejecutado por los periodistas, sino por los tecnólogos de cada equipo.

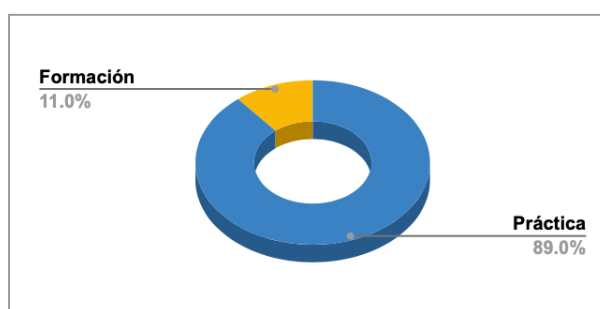
Si bien estas maneras de obtener información fueron importantes en ambos medios nativos digitales, la principal vía para la recopilación de datos consistió en la presentación de solicitudes de acceso a la información pública. De acuerdo a la periodista Milagros

Salazar¹⁹, la Ley, pese a sus debilidades, fue útil para que el equipo de *Convoca* acceda a más de mil informes ambientales que se encontraban encarpados y que de otra manera hubiera sido complicado recopilar. Además, detalló que el proceso para obtener información por medio de la Ley suele agotar a los periodistas y que no resulta rentable para un reportero de un medio tradicional —cuando podría estar produciendo notas periodísticas— dedicar tiempo a presentar cientos de solicitudes, hacerles seguimiento e insistir con las entidades públicas si la respuesta es inconforme.

A pesar de que en estos dos medios nativos digitales *Ojo Público* y *Convoca* es indispensable que los periodistas sepan redactar y presentar solicitudes de acceso a la información pública, no todos los reporteros tenían esos conocimientos cuando empezaron a laborar en los proyectos, ya que no había sido parte de su formación educativa superior. Tal como se muestra en la Figura 6, solo el 11 % de los periodistas encuestados para esta investigación aprendió el tema de acceso a la información pública en la formación universitaria, mientras que el resto lo hizo en la práctica periodística.

Figura 6

Adquisición de conocimientos de los periodistas sobre acceso a la información pública



¹⁹ Comunicación personal, 7 de setiembre de 2019.

Para el proyecto *Memoria robada* se presentaron 117 pedidos de información. El periodista David Hidalgo²⁰ recalca que, si bien en *Ojo Público* realizaron estas solicitudes al inicio de manera aleatoria para conocer más sobre el tema, cuando empezaron a recibir los documentos el equipo se percató de que tenía que realizar más pedidos debido a que la documentación recabada revelaba la existencia de otros datos que necesitaban, pero que no conocían que existían. Además, de acuerdo a su experiencia en el proyecto, la información que tiene el Estado en su poder suele estar desorganizada; es decir, se encuentra en diversas entidades y no está estructurada con una perspectiva de reúso, desde el punto de vista investigativo.

También recordó que la etapa de peticiones de información fue intensa debido a que muchas veces se tuvieron que cambiar las expresiones en la redacción para obtener información; como emplear las palabras “reportes” en vez de “informes”. Es decir, si no se usaba el término preciso, la entidad pública decía que esos documentos no existían. Para ello, se armó una estrategia periodística que tenía como objetivo conocer a profundidad los nombres de los documentos para pedirlos y evitar una respuesta negativa.

Por otro lado, en *Excesos sin castigo* se efectuaron cerca de 130 solicitudes de acceso a la información pública, principalmente al Organismo de Evaluación y Fiscalización Ambiental (OEFA). La cantidad de solicitudes fue alta porque varios requerimientos tuvieron que ser formulados nuevamente debido a que las respuestas no fueron satisfactorias.

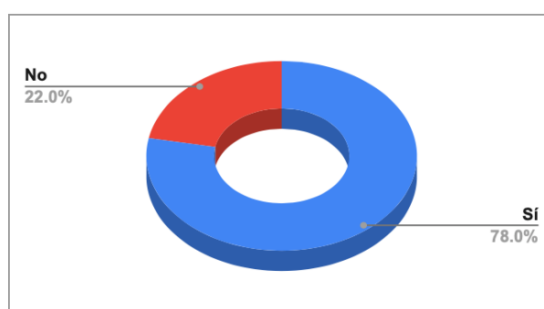
De acuerdo a la Figura 7, el 78 % de los periodistas encuestados de *Ojo Público* y *Convoca* presentó peticiones de acceso a la información. De ellos, al 86 % se le brindó respuestas fuera del plazo estipulado por Ley, información parcial, se le negó los datos, o se les consultó cómo usaría la información, pese a que tal cuestionamiento está prohibido

²⁰ Comunicación personal, 3 de junio de 2019.

en la normativa. La falta de una respuesta satisfactoria promovió la aplicación de recursos como la apelación administrativa o el hábeas data, pero solamente en un caso: el de *Memoria robada*.

Figura 7

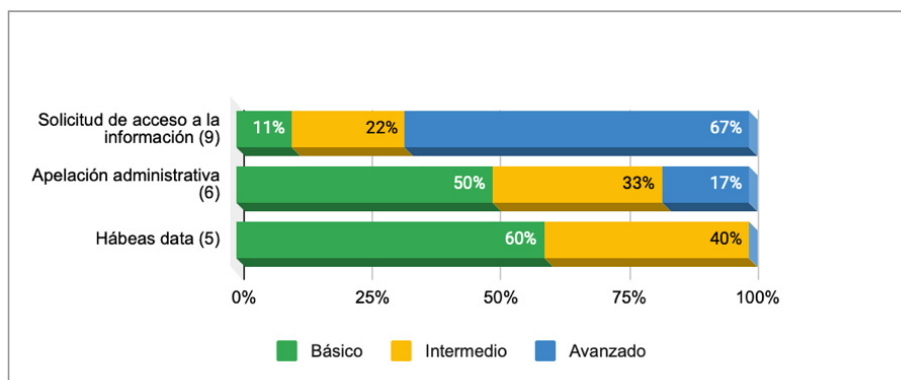
Periodistas que presentaron solicitudes de acceso a la información



De los nueve periodistas encuestados, solo el 67 % tenía conocimientos avanzados en la elaboración de solicitudes de acceso a la información pública; el 17 %, en la presentación de apelaciones administrativas, de un universo de seis reporteros, y ninguno de los cinco periodistas que aseguraron conocer el tema del hábeas data tenía habilidades avanzadas para redactar y presentar esta demanda, tal como se observa en la Figura 8. También, se aprecia que los niveles de habilidad de los periodistas descienden conforme se complican los procesos para obtener datos, como sucede con la apelación administrativa y el hábeas data.

Figura 8

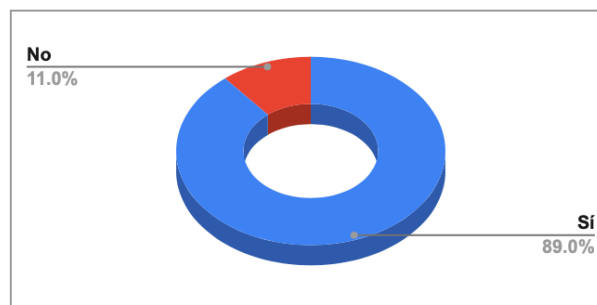
Trámites vinculados al proceso de acceso a la información pública conocidos por los periodistas y sus niveles de habilidad antes de iniciar el proyecto



Solo *Ojo Público* presentó una apelación administrativa y participó en un proceso de hábeas data durante la investigación. Este medio realizó una petición de información al Ministerio de Cultura para acceder a todas las copias simples de los certificados de exportación, emitidos entre enero de 2008 y diciembre de 2015, de aquellos bienes que no pertenecen al patrimonio cultural de la nación. Sin embargo, la documentación fue negada bajo el argumento de que contenía información protegida por el secreto comercial y que su publicación podría generar una ventaja competitiva a terceras personas.

5.1.2. Limpieza de datos

El 89 % de los periodistas encuestados participó en el procesamiento de datos, según se aprecia en la Figura 9. El grupo se encargó de transcribir la información más relevante a diferentes matrices en *Excel* con la finalidad de obtener bases de datos ordenadas que puedan ser usadas de acuerdo a los objetivos de cada equipo.

Figura 9*Periodistas que procesaron información*

Una de las características principales de los proyectos *Memoria robada* y *Excesos sin castigo* fue la construcción de bases de datos a partir de documentos técnicos que no se encontraban estructurados. Y, debido a que se disponía de mucha documentación para analizar y procesar, ambos equipos buscaron voluntarios para continuar con la labor. Al equipo de *Ojo Público* se sumaron dos estudiantes de periodismo y al de *Convoca*, siete. El objetivo fue terminar de construir las bases de datos en el menor tiempo posible; sin embargo, se detectaron errores en esta fase de trabajo.

Milagros Salazar²¹ señaló que más integrantes en el equipo no asegura calidad en el trabajo, sobre todo si no posee una vasta experiencia profesional: “un joven con poca experiencia no llena una base de datos como debería llenarla, por más capacitación que le des”. Por su parte, Fabiola Torres²², de *Ojo Público*, reveló que sus voluntarios se encargaron de procesar datos al inicio, pero debido a que “no había garantía de que fueran a llenar correctamente la información se les encargó construir un registro de imágenes”: una tarea completamente manual que no requirió analizar información.

Durante esta fase de procesamiento de datos, de los nueve periodistas encuestados, el 22 % empleó *Tabula*, una herramienta que permite extraer información de tablas en

²¹ Comunicación personal, 7 de setiembre de 2019.

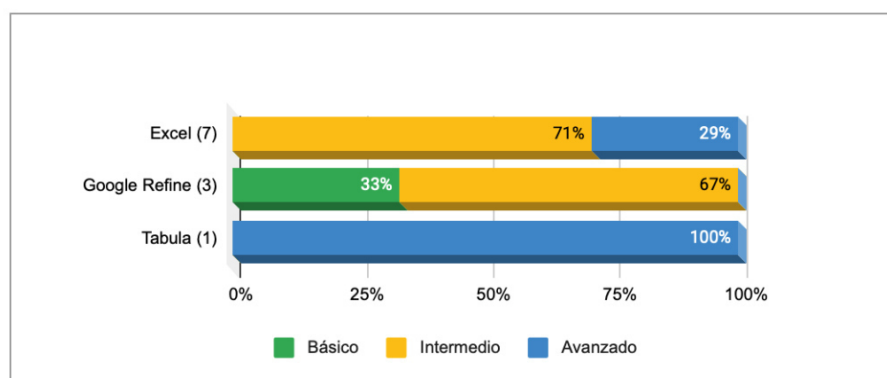
²² Comunicación personal, 6 de setiembre de 2019.

PDF y convertirlas en formatos más amigables como hojas de cálculo, con un nivel de conocimiento intermedio y avanzado. A su vez, el 22 % empleó *OCR*, un software que permite extraer textos dentro de imágenes, con un uso básico e intermedio, y el restante 56% empleó *Excel* con un nivel intermedio y avanzado. Ninguno mencionó haber utilizado otra herramienta.

Luego de completar las bases de datos se procedió a hacer la limpieza. El 78 % de periodistas; es decir siete, se involucró en esta fase. Para ello, el grupo empleó principalmente *Excel* y *Open Refine* con un nivel de conocimiento intermedio, y *Tabula*, en el caso de uno de los encuestados, con un uso avanzado, como se aprecia en la Figura 10. Conforme se requiere el empleo de más herramientas o programas en la labor periodística, se reduce la cantidad de personas capacitadas para usarlas.

Figura 10

Programas empleados por los periodistas y sus niveles de habilidad en la fase de limpieza de datos

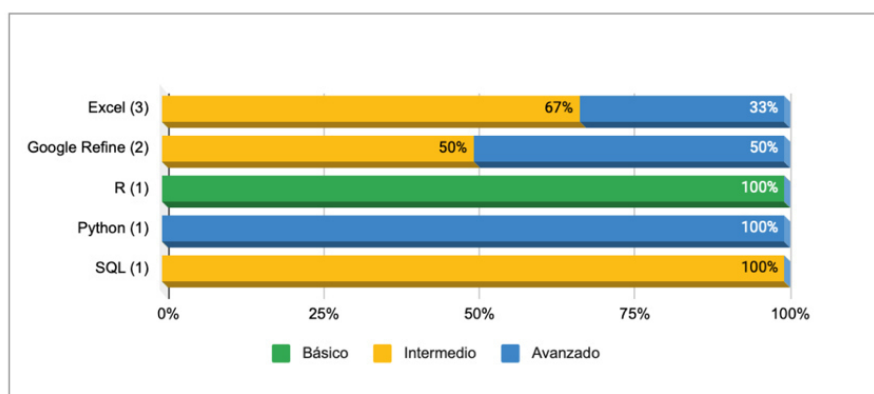


La limpieza de datos no solo consistió en uniformizar la información y eliminar errores ortográficos, sino en validar contenido que podría resultar errado como parte de un proceso de verificación inicial que se mantiene en toda la investigación.

En esta misma línea, los cuatro tecnólogos participaron en la limpieza de datos. Para este proceso utilizaron principalmente *Excel* y *Open Refine* con un nivel de conocimiento intermedio y avanzado, de acuerdo a la Figura 11. En menor medida emplearon los lenguajes de programación *R*, *SQL* y *Python*, con un nivel de uso básico, intermedio y avanzado, respectivamente.

Figura 11

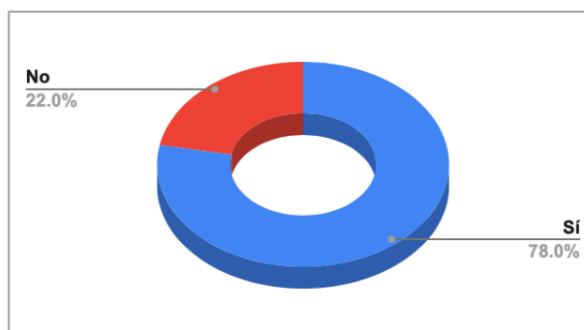
Programas y lenguajes empleados por los tecnólogos y su nivel de habilidad en la fase de limpieza de datos



Del total de periodistas encuestados, el 78 % participó en la estructuración de las bases de datos, de acuerdo a la Figura 12. Esto significa que definió los nombres de las columnas en las hojas de cálculo y el tipo de información que se colocaría en cada celda, para que luego los datos sean completados por el equipo de periodistas.

Figura 12

Participación de periodistas en la definición de las estructuras de las bases de datos



Según explicó Fabiola Torres²³, las matrices en *Memoria robada* se definieron en equipo tras varias reuniones, pero teniendo como base la propuesta inicial del periodista José Luis Huacles. Ella reconoció que no había una metodología establecida para esta y las demás fases de trabajo: “nuestra organización de los datos era muy instintiva, porque no partimos de ninguna metodología, y el trabajo lo fuimos afinando en el camino”.

En el caso de *Convoca*, las matrices también fueron estructuradas por periodistas: principalmente por Milagros Salazar y Esteban Valle Riestra, con la participación de Aramis Castro. A diferencia de *Convoca*, *Ojo Público* sí incluyó la labor de un tecnólogo para definir desde el inicio una matriz principal que posteriormente fue visualizada.

5.1.3. Contextualización y combinación de datos

En esta fase de trabajo se analizó el contexto de la información obtenida. Por ejemplo, todos los periodistas de ambos medios nativos digitales, a excepción de uno, conocían detalles de la información que recibieron en hojas de cálculo, como la fecha en que la información fue recabada por la entidad pública, la metodología de construcción de esas bases de datos, el significado de cada término incluido, el nombre del área de la entidad pública o de los responsables de recolectar la información, la fecha de recolección

²³ Comunicación personal, 6 de setiembre de 2019.

y las razones de la recolección. Además, coincidieron en que no todas las bases de datos recibidas de las entidades del Estado incluían un diccionario de datos que les permitiera conocer con facilidad todos los significados de los términos mencionados en las bases de datos o documentos recibidos para interpretarlos correctamente.

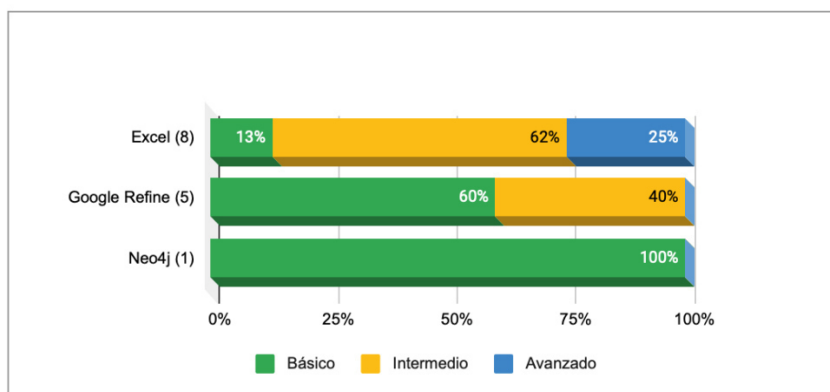
Luego de la contextualización de información se realizó la fase de combinación, que fue aplicada por los profesionales de ambos medios de comunicación. Sin embargo, un poco menos de la mitad participó del proceso: solo tres periodistas y un tecnólogo trabajaron en la combinación de entre dos a cuatro bases de datos, con la finalidad de hacer un cruce de información para ampliar, validar o comparar la data con información de otras bases de datos.

5.1.4. Análisis de datos

Los nueve periodistas encuestados participaron en el análisis de datos, pero con un nivel de involucramiento que varió según las dificultades que se presentaron en las tareas a realizar. Es decir, conforme se requirieron conocimientos de más programas o herramientas para el análisis de datos, el nivel de habilidad de los periodistas disminuyó, tal como se aprecia en la Figura 13. Por ejemplo, de los ocho periodistas que manejaron *Excel*; el 13 % lo hizo con un nivel básico, el 62 % con un nivel intermedio y el 25 % restante con uno avanzado. En esta línea, cinco reporteros, además de *Excel*, utilizaron *Open Refine (Google Refine)*, pero con un nivel básico e intermedio de 60 % y 40 %, respectivamente. Por último, solo uno de los encuestados aseveró haber empleado *Excel*, *Open Refine* y *Neo4j*; no obstante, el nivel de habilidad de este último fue básico.

Figura 13

Programas empleados por los periodistas y sus niveles de habilidad en la fase de análisis de datos



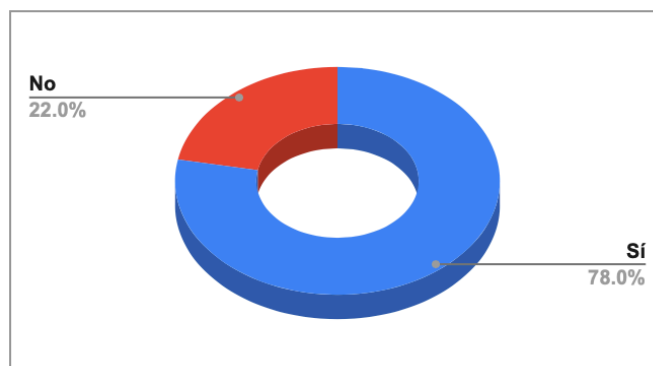
Es importante mencionar que el 75 % de los tecnólogos de ambos equipos participó en el análisis de datos. Para ello, empleó *Excel* a nivel intermedio, *Open Refine* con niveles de conocimiento básico y avanzado; *R* con habilidades básicas, *SQL* con conocimiento intermedio y *Python* con habilidad de uso avanzada.

5.1.5. Verificación de datos

Siete de los nueve periodistas encuestados; es decir, el 78 %, según se aprecia en la Figura 15, trabajaron en la verificación de datos. La tercera parte de todos los periodistas indicó que la verificación de la información contenida en las bases de datos fue parcial; es decir, que se aplicó solo a una parte de la información recopilada; mientras que el resto sostuvo que sí verificó toda la información.

Figura 14

Periodistas que participaron en la verificación de datos



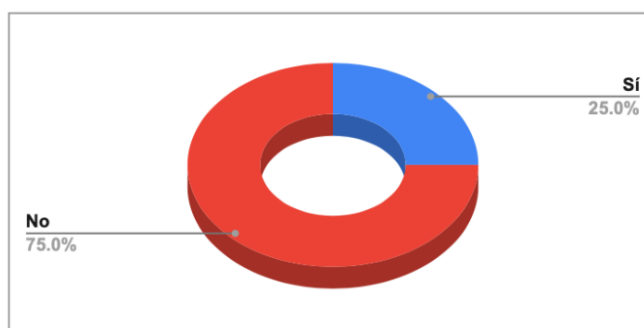
Sobre esta fase de trabajo, la periodista Milagros Salazar²⁴ manifestó que la verificación se realizó a toda la información que fue publicada en el sitio web de *Convoca* ya sea a través de gráficos, reportajes, cuadros, etc. Para ello, se realizaron entrevistas y solicitaron descargos a las personas u organismos involucrados en la investigación. Cabe mencionar que la información obtenida, al igual que en *Ojo Público*, provino de entidades del Estado; es decir, de fuentes confiables. No obstante, fue necesario validar los datos, porque dicha información es construida por personas que pueden cometer errores al momento de organizar su data.

En el proceso de verificación de datos participó solo uno de los tecnólogos encuestados, que perteneció a *Convoca*, tal como se aprecia en la Figura 15.

²⁴ Comunicación personal, 7 de setiembre de 2019.

Figura 15

Tecnólogos que participaron en la verificación de datos



En el caso de *Memoria robada*, la periodista Fabiola Torres²⁵ reconoció que no se validó el 100 % de la información recibida porque era abundante y por tratarse de registros públicos provenientes de fuentes confiables, como lo fue la información de todos los bienes registrados como patrimonio cultural del Perú. A pesar de ello, sí se validó todos los datos contenidos en las historias periodísticas que se publicaron al igual que lo hizo *Convoca*.

5.1.6. Comunicación de datos

En el equipo de *Convoca* todos los periodistas encuestados aseguraron haber elaborado artículos basados en datos; mientras que en *Ojo Público* fueron solamente dos periodistas debido que el resto del equipo se dedicó a otras actividades, como el procesamiento de información, análisis de datos y elaboración de materiales para la difusión de las investigaciones en la web del medio y en sus redes sociales. Esto no impidió que los periodistas colaboraran con datos puntuales que nutrieron los reportajes.

El 56 % de los nueve periodistas encuestados en los medios nativos digitales, además de los reportajes, participó en la elaboración de conjuntos de datos abiertos, visualizaciones interactivas e infografías, y material para difusión del proyecto. Para ello,

²⁵ Comunicación personal, 6 de setiembre de 2019.

emplearon programas como *Adobe Premier*, *Infogram* y *Tableau*.

En tanto, el 89 % de los periodistas de *Ojo Público* y *Convoca* participó en la difusión *online* del proyecto con la elaboración de publicaciones para redes sociales. De ellos, el 63 % también realizó notas de prensa, videos, gráficos para redes sociales o *Newsletter*.

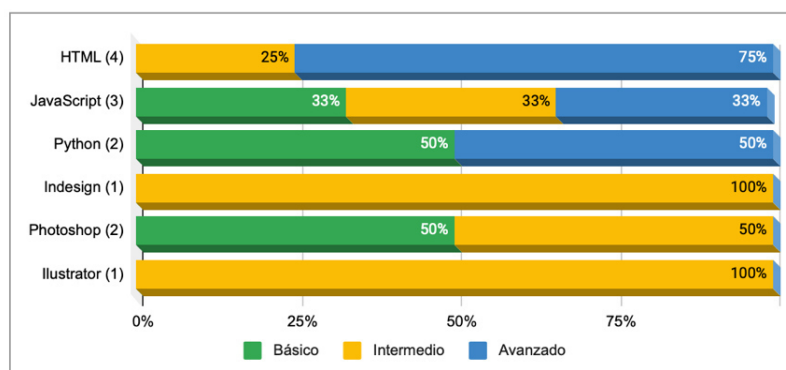
Por otro lado, solo el 44 % de todos los periodistas participó en la estrategia *offline* de difusión. Esta consistió en entrevistas, conferencias y talleres. Es importante destacar que usualmente los coordinadores y editores de proyectos son los que tienen más actividad de difusión *offline* debido a que conocen a profundidad los detalles periodísticos de las investigaciones, cómo surgió la idea para elaborarlas y se fue ejecutando.

Los tecnólogos también tuvieron un rol importante en la difusión de la información. Por ejemplo, el de *Ojo Público* hizo toda la labor de diseño e implementación de visualizaciones y el sitio web, con la recepción de materiales gráficos y audiovisuales que el medio solicitó a una diseñadora y a dos profesionales audiovisuales. Mientras que en *Convoca* la tarea fue distribuida entre tres personas.

De acuerdo la Figura 16, para la tarea de diseño e implementación de los sitios web, los cuatro tecnólogos de ambos medios usaron *HTML*, un lenguaje de marcado de texto: el 25 % con un nivel de conocimiento intermedio y el 75 %, avanzado. Tres utilizaron el lenguaje de programación *JavaScript* y dos, *Python*. Solo uno de los cuatro encuestados afirmó haber usado los programas *Indesign* e *Illustrator*. Estos programas son utilizados principalmente por el equipo de periodistas.

Figura 16

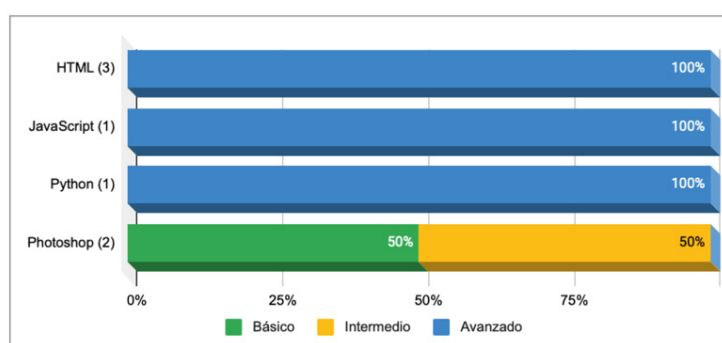
Programas y lenguajes empleados por los tecnólogos en el diseño e implementación de los sitios web



Por otro lado, la Figura 17 muestra los programas, lenguaje de marcado de texto (*HTML*) y lenguajes de programación empleados en el diseño e implementación de visualizaciones.

Figura 17

Programas y lenguajes empleados por los tecnólogos en el diseño e implementación de visualizaciones



De los cuatro tecnólogos encuestados, tres aseguraron haber empleado *HTML* con un nivel de conocimiento avanzado; mientras que solo uno usó los lenguajes de programación *JavaScript* y *Python*.

DISCUSIÓN

Los resultados evidencian que en los proyectos periodísticos *Memoria robada*, de *Ojo Público*, y *Excesos sin castigo*, de *Convoca*, los periodistas y tecnólogos sí siguieron la metodología de trabajo del periodismo de datos, que para esta investigación se estructuró en base a los conceptos de Bradshaw (2011) y Segnini (2012, como se citó en Ramírez, 2012). Sin embargo, esta metodología no fue estipulada desde el inicio por lo que fue perfeccionándose e implementándose con el transcurso de las investigaciones; puesto que los periodistas no contaban con un plan de trabajo riguroso para cada fase de trabajo ni tampoco tenían una vasta experiencia en periodismo de datos.

Además, los conocimientos que los periodistas poseían en el uso de herramientas tecnológicas que les permitieron desarrollar sus labores eran básicos e intermedios en la mayoría de casos. Por ejemplo, de los ocho reporteros que aseguraron emplear *Excel* en el análisis de datos, el 13 % comunicó que lo manejaba de manera básica, el 62 % en un nivel intermedio y el 25 % tenía una habilidad de uso avanzada. Conforme se requirió el manejo de otras herramientas, como *Open Refine* o *Neo4j*, el nivel que predominaba en sus usos era el básico. Una situación similar se observó en las fases de recolección, limpieza y combinación de datos.

De la situación anterior se desprende que los reporteros podían resolver cuestiones básicas en la limpieza, análisis y combinación de datos, pero cuando requerían de conocimientos más avanzados las tareas eran compartidas con los tecnólogos, en ambos proyectos periodísticos. Esto confirma la necesidad de que los equipos profesionales del periodismo de datos sean multidisciplinarios (Bueno, 2014).

Estos últimos profesionales, los tecnólogos, por el contrario, sí manejaron herramientas tecnológicas y lenguajes de programación con un nivel intermedio y avanzado en la mayoría de casos. Debido a que se requerían ciertas habilidades para la fase

de comunicación de datos, solo un poco más de la mitad de los periodistas (56 %) participó en esa etapa del trabajo.

La investigación también muestra que, durante el desarrollo de los proyectos periodísticos, en la fase de recopilación de información a través de la Ley de Transparencia y Acceso a la Información Pública, los periodistas se enfrentaron a respuestas negativas por parte de las entidades públicas, uno de los tantos problemas identificados en la aplicación de la Ley (Ministerio de Justicia y Derechos Humanos, 2022). No obstante, solo un medio realizó una apelación administrativa y luego un hábeas data.

Desde el lado del acceso a la información pública, también se evidenció que los periodistas participantes de los proyectos *Memoria robada* y *Excesos sin castigo* tenían conocimientos para realizar peticiones a las instituciones públicas: 11 % con un nivel de habilidad básico; el 22 %, intermedio y el 67 %, avanzado. Sin embargo, no manejaron los mismos niveles de habilidades en las apelaciones administrativas y en la elaboración de demandas de hábeas data. Para proyectos periodísticos de este tipo, donde la fuente principal es información recabada gracias a la Ley, resulta indispensable que estén capacitados en cómo responder ante la negativa de información, tanto en la vía administrativa como judicial.

Eso nos lleva a reflexionar sobre la situación de la formación académica de pregrado en universidades peruanas. No todas las casas de estudio dictan cursos, que no sean electivos, donde se profundice sobre la legislación en materia de transparencia y acceso a la información. Asimismo, en la línea de Cárdenas (2020), los funcionarios suelen interpretar erróneamente los alcances de la normativa, como el caso del hábeas data que interpuso *Ojo Público* contra el Ministerio de Cultura. Esta entidad pública argumentó que la información solicitada estaba protegida por el secreto comercial; no obstante, el caso fue resuelto a favor del medio nativo digital el 2021.

El apoyo de los tecnólogos en la recopilación de datos, a través del *scraping*, y el análisis, limpieza, combinación y comunicación de información fue relevante para acelerar los procesos de trabajo y culminar las actividades. Pese a resultar una pieza fundamental en el desarrollo del periodismo de datos, no siempre se involucran en el trabajo periodístico desde la primera fase de trabajo que es la recopilación de información. Su participación durante todo el proceso sería de utilidad para que el equipo pueda planear, en conjunto, que tipo de información complementaria solicitar y, en base a los avances, pensar en la manera cómo se va a comunicar la información a través de los productos periodísticos.

CONCLUSIONES

En esta investigación se ha llegado a las siguientes conclusiones:

1. Los periodistas y tecnólogos que participaron en la elaboración de los proyectos *Memoria robada* y *Excesos sin castigo*, de *Ojo Público* y *Convoca*, respectivamente, desarrollaron todas las fases del proceso metodológico del periodismo de datos: recopilación, limpieza, contextualización, combinación, análisis, verificación y comunicación de información, refutándose así la hipótesis planteada. Sin embargo, gran parte de los conocimientos aplicados era de nivel básico, tanto en el uso de programas y herramientas tecnológicas que se requirieron emplear en cada fase de trabajo, así como en la aplicación de la Ley.
2. El proceso de recopilación de información en los proyectos periodísticos *Memoria robada* y *Excesos sin castigo* incluyó la realización de entrevistas, extracción de información de portales web y, en mayor cantidad, peticiones de acceso a la información pública. Respecto a este último procedimiento, se observó que los periodistas que participaron en las investigaciones tenían conocimientos de cómo acceder a información; sin embargo, ese conocimiento fue básico al momento de apelar por la vía administrativa o judicial las respuestas negativas de brindar información de las entidades públicas.
3. Para la fase de limpieza de información se requería a periodistas con conocimientos intermedios y avanzados de herramientas tecnológicas como por ejemplo *Excel* y *Open Refine* con la finalidad de acelerar los procesos de trabajo y también para que la limpieza no recaiga sobre un pequeño grupo de reporteros.
4. Todos los periodistas que participaron en los proyectos de *Ojo Público* y *Convoca* contextualizaron los datos obtenidos, de documentos o bases de datos, para las investigaciones realizadas; es decir, conocían su procedencia, el significado de los términos, el área o personas encargadas de las entidades públicas de crear dicha información. No obstante, ello requirió, en muchos casos, averiguaciones de los reporteros,

ya que no todas las bases de datos contenían un diccionario de datos que les permitiera comprender la información contenida en ellas.

5. Según los periodistas encuestados, la creación de una o varias matrices, en la fase de combinación de datos, permitió comprender mucho mejor la información para encontrar relaciones. Estos hallazgos no solo fueron empleados en los reportajes, sino también en las visualizaciones interactivas.
6. A pesar de que se requerían conocimientos básicos de lenguajes de programación en la fase de análisis de datos, cuando se manejaron grandes volúmenes de información, los periodistas no estaban formados con ese tipo de conocimiento. Optaron, en cambio, por el empleo del programa *Excel*, la aplicación de escritorio *Open Refine* y el software *Neo4j*, aunque el nivel de manejo de estos dos últimos fue principalmente básico. En este proceso del periodismo de datos fue importante la participación de los tecnólogos, quienes, a través lenguajes de programación, culminaron la fase del análisis de información.
7. Los equipos periodísticos de *Ojo Público* y *Convoca* verificaron la información recabada para las investigaciones de manera parcial debido a que se trataban de miles de datos que provenían de fuentes oficiales, y que humanamente era imposible contrastar. En ambos medios, la información que se utilizó en los reportajes, cuadros, gráficos y otras piezas periodísticas sí fue verificada por medio de consultas a las entidades públicas y validaciones con otras fuentes; a excepción de la información contenida en los buscadores de *Ojo Público* que no fue validada por tratarse de millones de datos.
8. Por el lado de los reporteros de los medios nativos digitales en mención, la comunicación de datos consistió en la redacción de reportajes, y piezas informativas que se difundieron de manera *offline* y en las redes sociales con la ayuda de programas de edición como *Adobe Premier* y *Photoshop*; así como visualizaciones interactivas a través de programas gratuitos como *Tableau* e *Infogram*. Mientras que los tecnólogos, para la creación de

mapas, visualizaciones o buscadores, emplearon *HTML* y lenguajes de programación como *JavaScript* y *Python*; y en algunos casos programas de diseño como *Indesign* e *Illustrator*.

RECOMENDACIONES

Como consecuencia de la investigación se recomienda lo siguiente:

1. A las universidades se les aconseja incluir en los programas de pregrado de formación periodística de las universidades peruanas cursos obligatorios donde se enseñe el periodismo de datos; así como los conocimientos y herramientas necesarias para su aplicación, como lenguajes de programación, *Excel*, estadística y acceso a la información pública. Si bien algunas universidades mencionan la Ley de Transparencia y Acceso a la Información Pública en sus clases, esta suele ser parte de cursos electivos y en los que no se profundizan todos los aspectos de la Ley.
2. Las instituciones públicas no solo deben responder las solicitudes de acceso a la información dentro del plazo estipulado en la Ley, sino entregar los datos en formatos amigables que puedan ser reutilizados, como lo son las hojas de cálculo o archivos *CSV*.
3. A los periodistas se les sugiere apelar por la vía administrativa aquellas solicitudes de acceso a la información que fueron negadas por parte de las entidades públicas. Si el Tribunal de Acceso a la Información Pública, que es la última instancia administrativa en estos temas, no falla a su favor y considerando que tienen la razón, pueden llevar el caso al Poder Judicial mediante un hábeas data.
4. A las organizaciones periodísticas se les recomienda que, además de los periodistas y tecnólogos, los diseñadores también deben participar en la fase de comunicación de datos. Esto con la finalidad de que la presentación de los productos, ya sean reportajes, buscadores o bases de datos, sean visualmente amigables con el lector.
5. Con el fin de promover la transparencia, se recomienda, a los medios nativos digitales que desarrollan investigaciones con datos masivos, detallar el proceso metodológico que siguieron. Asimismo, liberar en formatos reutilizables las bases de datos que emplearon para que sean utilizadas por otros periodistas, estudiantes, académicos, etc.

REFERENCIAS BIBLIOGRÁFICAS

- Blanco, E., Teruel, L. & Martín, V. (2016). El periodismo de datos en España. Más calidad y valor añadido a la información. *Cuadernos Artesanos de Comunicación* 112. *Periodismo de datos*. 13-37.
<https://issuu.com/revistalatinadecomunicacion/docs/cac112>
- Bradshaw, P. (2011). The inverted pyramid of data journalism. *Online Journalism Blog*.
<https://onlinejournalismblog.com/2011/07/07/the-inverted-pyramid-of-data-journalism/>
- Bueno, A. (2014). *Periodismo de Datos. Antecedentes, ámbitos de desarrollo y aplicación en medios generalistas: el caso de El Confidencial y El Periódico*. [Trabajo de Fin de Grado en Periodismo]. Universidad de Valladolid.
https://uvadoc.uva.es/bitstream/handle/10324/5752/TFG_F_2014_8.pdf;jsessionid=6C9DDAD44B62EFF7DD3ED47C187F3A33?sequence=1
- Cárdenas, H. (2020). *Deficiencias y limitaciones normativas al derecho de acceso a la información pública en el Perú*. [Tesis de Doctorado]. Universidad Nacional Mayor de San Marcos. <https://hdl.handle.net/20.500.12672/15612>
- Cortés, H., Luengo, M. & Elías, C. (2018). Periodismo de datos y transparencia al margen de los grandes medios: Un estudio comparativo de Civio y ProPublica. *Icono*. Volumen 16, número 2, 66-87.
<https://icono14.net/ojs/index.php/icono14/article/view/1177/1305>
- Crucianelli, S. (2013). 12 puntos para entender el periodismo de datos. *Knight Center for Journalism in the Americas*. The University of Texas at Austin.
<https://knightcenter.utexas.edu/es/blog/00-13717-12-puntos-para-entender-el-periodismo-de-datos>
- Crucianelli, S. (2013b). ¿Qué es el periodismo de datos? *Cuadernos de periodistas*. Volumen 26, 106-124.
http://www.cuadernosdeperiodistas.com/pdf/Cuadernos_de_Periodistas_26.pdf
- Decreto Supremo N° 021-2019-JUS. (11 de diciembre de 2019). Normas Legales. Diario Oficial El Peruano. <https://busquedas.elperuano.pe/download/url/texto-unico-ordenado-de-la-ley-n-27806-ley-de-transparenci-decreto-supremo-n-021-2019-jus-1835794-3>
- Decreto Supremo N° 072-2003-PCM. (7 de agosto de 2003). Normas Legales. Diario Oficial El Peruano. <https://www.peru.gob.pe/normas/docs/DS-072-2003-PCM.pdf>
- Esteban, Ch. (2012). Las nuevas profesiones del periodismo. *Cuadernos Evoca*. Volumen 7, 17-23. <http://www.evocaimagen.com/cuadernos/cuadernos7.pdf>
- Fallas, H. (2018). Las seis columnas del periodismo de datos. *La Data Cuenta*.
<http://hasselfallas.com/2016/05/03/las-seis-columnas-del-periodismo-de-datos/>

- Fallas, H. (2014). Entrevista con una base de datos. *Manual de Periodismo de Datos Iberoamericano*. <http://manual.periodismodedatos.org/hassel-fallas.php>
- Ferreras, E. (2013). Aproximación teórica al perfil profesional del Periodista de datos. *Ícono 14*, volumen 11 (2), 115-140. doi: 10.7195/ri14.v11i2.573
<https://dialnet.unirioja.es/descarga/articulo/4721933.pdf>
- Flores, J. & Salinas, C. (2013). El periodismo de datos como especialización de las organizaciones de noticias en Internet. *Correspondencias & Análisis*, N° 3.
<https://dialnet.unirioja.es/descarga/articulo/4739290.pdf>
- Forero, M. (2013). *Los primeros pasos del periodismo de bases de datos en Colombia. Estudios de caso de El Espectador y La Silla Vacía*. [Trabajo de Grado Profesional]. Fundación Universitaria INPAHU. <https://labmedia.org/wp-content/uploads/docs/ES/Los%20primeros%20pasos%20de%20periodismo%20de%20bases%20de%20datos%20en%20Colombia.pdf>
- García, A. & Catalina, B. (2018). Una perspectiva documental y bibliotecológica sobre el big data y el periodismo de datos. *Investigación Bibliotecológica*. Volumen 32, número 74, 77-99. <http://rev-ib.unam.mx/ib/index.php/ib/article/download/57910/51875>
- Gordon, R. (2013). Want to build a data journalism team? You'll need these three people. *Knight lab*. <https://knightlab.northwestern.edu/2013/06/28/want-to-build-a-data-journalism-team-youll-need-these-three-people/>
- Hidalgo, D. & Torres, F. (2016). *La navaja suiza del reportero*. Perú: Ojo Público y Consejo de la Prensa Peruana.
- Instituto Prensa y Sociedad de Venezuela (2010). *Métodos de la impertinencia*. <https://ipysvenezuela.org/2016/11/30/metodos-de-la-impertinencia-2/>
- Kvale, S. (2008). *Las entrevistas en investigación cualitativa*. Madrid: Ediciones Morata.
- Lee, M. (2014). *La investigación a partir de historias. Manual para periodistas de investigación*. Unesco. <http://unesdoc.unesco.org/images/0022/002264/226457S.pdf>
- Ley N° 29733. Ley de Protección de Datos Personales. (3 de julio de 2011). <https://diariooficial.elperuano.pe/pdf/0036/ley-proteccion-datos-personales.pdf>
- López, P. (2016). Capítulo I: Antecedentes y fundamentos del periodismo de datos. *Periodismo de datos. Nuevas narrativas para el Periodismo Especializado*, 5-26. <https://riuma.uma.es/xmlui/bitstream/handle/10630/12465/Libro%20de%20actas%20RIUMA%20Cap1.pdf?sequence=3>
- López, P. (2019). *La enseñanza-aprendizaje del periodismo de datos en las facultades españolas. Estudios y propuestas de viabilidad*. [Tesis Doctoral]. Universidad Complutense de Madrid. <https://eprints.ucm.es/id/eprint/50688/1/T40749.pdf>

- Magallón, R. (2013). *¿Por qué es importante una buena ley de transparencia para el periodismo y para los periodistas?* Instituto para la Innovación Periodística.
http://www.academia.edu/5251182/_Por_qu%C3%A9_es_importante_una_buena_ley_de_transparencia_para_el_periodismo_y_para_los_periodistas_
- Mariño, Á. (2012). *La relación del periodismo de datos con big data y open data*. [Trabajo de Máster]. Universidad Carlos III de Madrid.
https://issuu.com/clog_unlp/docs/periodismodatosbigdataopendata
- Meyer, P. (1993). *Periodismo de precisión. Nuevas fronteras para la investigación periodística*. Bosch.
- Ministerio de Justicia y Derechos Humanos (2022). *Informe anual sobre solicitudes de acceso a la información pública*.
<https://cdn.www.gob.pe/uploads/document/file/3103438/Informe%20Anual%202021.pdf>
- Perla, J. (2010). El derecho de acceso a la información pública y el Hábeas Data en el Perú. *Diálogos de la Comunicación, número 82*.
https://repositorio.ulima.edu.pe/bitstream/handle/20.500.12724/2602/Perla_Anaya_Jose.pdf?sequence=1&isAllowed=y
- Proética (2009). *Periodismo de investigación: una guía práctica*.
<https://www.corteidh.or.cr/tablas/CD0387-6.pdf>
- Profeta, D. (2018). El periodismo de datos como oportunidad para renovar el periodismo, su desarrollo en América Latina y las competencias profesionales necesarias para su expansión. [Tesis de Magíster]. Universidad San Andrés.
<https://repositorio.udesar.edu.ar/jspui/bitstream/10908/18101/1/%5BP%5D%5BW%5D%20T.M.%20Per%20Profeta%20Molinuevo,%20Damián%20Ariel.pdf>
- Quiroz, R. (2016). El hábeas data, protección al derecho a la información y a la autodeterminación informativa. *Revista Letras 87(126)*.
http://www.scielo.org.pe/scielo.php?script=sci_arttext&pid=S2071-50722016000200002
- Quispe, C. (2016). *El acceso a datos como parte del derecho de acceso a la información pública*. [Tesis de Licenciatura]. Pontificia Universidad Católica del Perú.
http://tesis.pucp.edu.pe/repositorio/bitstream/handle/20.500.12404/6704/QUISPE_VALENCIA_CINDY_ACCESO_DATOS.pdf?sequence=8&isAllowed=y
- Ramírez, F. (2012). *Periodismo de datos, periodismo de siempre*.
<https://www.lapatria.com/blogs/periodismo-de-datos-periodismo-de-siempre>
- Ramonet, I. (2011). *La explosión del periodismo. Internet pone en jaque a los medios tradicionales*. Argentina: Capital Intelectual.
https://profesorsergiogarcia.files.wordpress.com/2016/03/la_explasion_del_periodismo_de_ramonet.pdf

- Ríos, M. (2014). *Open Refine – qué es + tutorial*. Escuela de datos. <https://bit.ly/3sxND0v>
- Salamanca, N., Sierra, J. & Huertas, C. (2018) *El periodismo de investigación latinoamericano en la era digital*. International Center for Journalists (ICFJ) y Connectas. http://www.connectas.org/RECURSOS/manual_de_periodismo_ICFJ-CONNECTAS.compressed.pdf
- Segnini, G. (2008). *Introducción al Periodismo Investigativo y al Periodismo Asistido por Computadora*. USAID. <https://desocultar.files.wordpress.com/2010/04/segnini-introduccion-al-periodismo-investigativo-y-al-periodismo-asistido-por-computadora.pdf>
- Spagnuolo, S. (2019). Cómo organizar un equipo de periodismo de datos. *Red Internacional de Periodistas*. <https://ijnet.org/es/story/c%C3%B3mo-organizar-un-equipo-de-periodismo-de-datos>
- Tuñas, E. (2015). *El periodismo de datos en 2015: análisis de El Mundo y The Guardian*. [Trabajo de Fin de Grado en Periodismo]. Universidad de Santiago de Compostella. https://www.academia.edu/17303850/Periodismo_de_Datos_en_2015_análisis_de_El_Mundo_y_The_Guardian
- Weiss, J. (2013). Tabula, una herramienta para agilizar el periodismo de datos. *Red Internacional de Periodistas*. <https://ijnet.org/es/story/tabula-una-herramienta-para-agilizar-el-periodismo-de-datos>

ANEXOS

Anexo 1: Cuestionario dirigido a periodistas que participaron en el proyecto *Excesos sin castigo*, de *Convoca*, y *Memoria robada*, de *Ojo Público*

Gracias por completar este cuestionario. Las respuestas son anónimas y serán empleadas solo con fines académicos. Por favor, llene la información solicitada y marque con un aspa (X), según corresponda. **Considere que las respuestas corresponden al periodo que trabajó únicamente para el proyecto *Excesos sin castigo* () o *Memoria robada* ().**

Datos generales

1. Género:

- () Masculino
() Femenino
() Otro: _____

2. ¿Cuántos años tenía cuando empezó a trabajar en la investigación periodística *Excesos sin castigo* o *Memoria robada*?

3. En ese momento, ¿qué nivel de estudios había alcanzado?

- () Estudiante
() Bachiller
() Licenciatura
() Maestría
() Doctorado

4. Especifique la carrera profesional:

5. Indique su centro de estudios:

6. ¿Qué cargo desempeñó en el proyecto *Excesos sin castigo* o *Memoria robada*?

7. ¿Cuántos años de experiencia laboral tenía en periodismo de datos al momento de iniciar el proyecto *Excesos sin castigo* o *Memoria robada*?

- () Menos de 1 año
() De 1 a 3 años
() De 4 a 7 años
() Más de 8 años
() No poseía experiencia

Recopilación de datos

8. ¿Adquirió información en virtud de la Ley de Transparencia y Acceso a la Información Pública para el proyecto *Excesos sin castigo o Memoria robada*? Si la respuesta es negativa, marque y pase a la pregunta 12.

- Sí
 No

9. ¿Qué inconvenientes experimentó al acceder a información pública para la investigación *Excesos sin castigo o Memoria robada*? Puede marcar más de una opción.

- Mala atención en mesa de partes
 Problemas técnicos con el formulario virtual
 Falta de respuesta
 Demora en la respuesta
 Información denegada
 Entrega de información parcial
 Cobros indebidos
 Consultas sobre cómo usará los datos
 No presenté inconvenientes
 Otro: _____

10. ¿Presentó alguna apelación administrativa?

- Sí
 No

11. ¿Presentó un hábeas data?

- Sí
 No

12. Marque los trámites vinculados al proceso de solicitud de información pública que conocía al iniciar el proyecto y su nivel de habilidad para elaborarlos: B (básico), I (intermedio) o A (avanzado). Puede marcar más de una opción.

- Solicitudes de acceso a la información pública:
• Nivel: B () I () A ()
 Cartas de apelación:
• Nivel: B () I () A ()
 Hábeas data:
• Nivel: B () I () A ()

13. ¿De qué manera adquirió sus primeros conocimientos sobre transparencia y acceso a la información pública?

- En la práctica periodística
 En la formación universitaria
 Capacitaciones
 Otro: _____

14. ¿Qué otras vías de recolección de información aplicó en el proyecto *Excesos sin castigo o Memoria robada*? Puede marcar más de una opción.

- () Búsqueda avanzada en sitio web
 () Entrevistas
 () *Scraping*
 () Otro: _____

Limpieza de datos

15. ¿Participó en la limpieza de datos? Si su respuesta es negativa, marque y pase a la pregunta número 17.

- () Sí
 () No

16. Marque las herramientas que empleó para la limpieza de datos y el nivel de habilidad que poseía: B (básico), I (intermedio) o A (avanzado). Puede marcar más de una opción.

- () Excel:
 • Nivel: B () I () A ()
 () Open Refine:
 • Nivel: B () I () A ()
 () Tabula:
 • Nivel: B () I () A ()
 () Otro (especifique el nivel): _____

17. ¿Participó en la definición de estructuras de las bases de datos construidas para la investigación *Excesos sin castigo o Memoria robada*?

- () Sí
 () No

18. ¿Transcribió información de formato físico y/o virtual a una hoja de cálculo? Si su respuesta es negativa, marque y pase a la pregunta número 20.

- () Sí
 () No

19. Marque las herramientas y/o programas tecnológicos que usó en la labor de transcripción de información y complete su nivel de habilidad: B (básico), I (intermedio) o A (avanzado). Puede marcar más de una opción.

- () Tabula:
 • Nivel: B () I () A ()
 () OCR:
 • Nivel: B () I () A ()
 () Excel:
 • Nivel: B () I () A ()
 () Todo lo efectué manualmente
 () Otro (especifique el nivel): _____

Contextualización de datos

20. ¿Qué aspectos conocía sobre las bases de datos recibidas de las instituciones públicas? Puede marcar más de una opción.

- () Responsable de recolectar la información
 () Fecha de recolección
 () Razones de la recolección
 () Metodología de construcción
 () Significado de cada término
 () Ninguna de las anteriores

21. ¿Todas las bases de datos recibidas de las entidades públicas incluían un diccionario de datos?

- () Sí
 () No
 () Desconozco

Combinación de datos

22. ¿Participó en la combinación de bases de datos? Si la respuesta es negativa, marque y pase a la pregunta número 24.

- () Sí
 () No

23. ¿Cuántas bases de datos combinó?

- () 2
 () 3
 () 4
 () Otro: _____

Análisis de datos

24. ¿Participó en la fase de análisis de datos? Si la respuesta es negativa, marque y pase a la pregunta número 26.

- () Sí
 () No

25. Marque los programas y/o herramientas que empleó para el análisis de datos en el proyecto *Memoria robada* o *Excesos sin castigo* y complete su nivel de habilidad: B (básico), I (intermedio) o A (avanzado). Puede marcar más de una opción.

- () Excel:
 • Nivel B () I () A ()
 () Open Refine:
 • Nivel B () I () A ()
 () R:
 • Nivel B () I () A ()
 () Otro (especifique el nivel): _____

Verificación de datos

26. ¿Participó en la verificación de la información contenida en las bases de datos de la investigación *Excesos sin castigo o Memoria robada*? Si la respuesta es negativa, marque pase a la pregunta 28.

- Sí
 No

27. ¿Cómo fue el proceso de verificación de las bases de datos?

- Total
 Parcial

Comunicación de datos

28. ¿Cuál de los siguientes productos periodísticos elaboró? Si no marcó la opción “visualizaciones interactivas” y/o marcó “ninguno”, pase a la pregunta número 30. Puede marcar más de una opción.

- Artículos basados en datos
 Visualizaciones interactivas
 Conjunto de datos abiertos
 Aplicaciones de noticias
 Infografías
 Videos
 Audios
 Ninguno
 Otro: _____

29. ¿Qué programas/herramientas empleó para la creación de visualizaciones interactivas en el proyecto *Excesos sin castigo o Memoria robada*? Puede marcar más de una opción.

- Infogram
 Datawrapper
 Google Fusion Tables
 HTML
 Otro: _____

30. ¿Qué tipo de materiales periodísticos elaboró para la difusión *online* del proyecto *Excesos sin castigo o Memoria robada*? Puede marcar más de una opción.

- Post para redes sociales
 Gráficos para redes sociales
 Newsletter
 Notas de prensa
 No participé en la estrategia *online*
 Otro: _____

31. ¿En qué estrategias *offline* de difusión del proyecto periodístico *Excesos sin castigo o Memoria robada* participó? Puede marcar más de una opción.

- () Entrevistas
- () Conferencias
- () Talleres
- () No participé en la estrategia *offline*
- () Otro: _____

¡Gracias por responder este cuestionario!

Anexo 2: Cuestionario dirigido a tecnólogos que participaron en el proyecto *Excesos sin castigo*, de *Convoca*, y *Memoria robada*, de *Ojo Público*

Gracias por completar este cuestionario. Las respuestas son anónimas y serán empleadas solo con fines académicos. Por favor, llene la información solicitada y marque con un aspa (X), según corresponda. **Considere que las respuestas corresponden al periodo que trabajó únicamente para el proyecto *Excesos sin castigo* () o *Memoria robada* ().**

Datos generales

1. Género:

- () Masculino
() Femenino
() Otro: _____

2. ¿Cuántos años tenía cuando empezó a trabajar en la investigación periodística *Excesos sin castigo* o *Memoria robada*?

3. En ese momento, ¿qué nivel de estudios había alcanzado?

- () Estudiante
() Bachiller
() Licenciatura
() Maestría
() Doctorado

4. Especifique la carrera profesional:

5. Indique su centro de estudios:

6. ¿Qué cargo desempeñó en el proyecto *Excesos sin castigo* o *Memoria robada*?

7. ¿Cuánto tiempo de experiencia laboral tenía como programador al momento de iniciar el proyecto periodístico *Excesos sin castigo* o *Memoria robada*?

- () Menos de 1 año
() De 1 a 3 años
() De 4 a 7 años
() Más de 8 años
() No poseía experiencia

8. ¿Tenía conocimientos sobre periodismo de datos al iniciar el proyecto?

- () Sí
() No

Recopilación de datos

9. ¿Qué métodos de recopilación de información aplicó en el proyecto *Excesos sin castigo o Memoria robada*?

- () Scraping
 () No recopilé datos
 () Otro: _____

Limpieza de datos

10. ¿Participó en la limpieza de datos de la investigación *Excesos sin castigo o Memoria robada*? Si su respuesta es negativa, marque y pase a la pregunta número 12.

- () Sí
 () No

11. Marque las herramientas y/o lenguajes de programación que empleó para la limpieza de datos y complete su nivel de habilidad: B (básico), I (intermedio) o A (avanzado). Puede marcar más de una opción.

- () Excel:
 • Nivel B () I () A ()
 () Open Refine:
 • Nivel B () I () A ()
 () R:
 • Nivel B () I () A ()
 () Otro (especifique el nivel): _____

12. ¿Participó en la definición de estructuras de las bases de datos construidas para la investigación *Memoria robada o Excesos sin castigo*?

- () Sí
 () No

Combinación de datos

13. ¿Participó en la combinación de bases de datos del proyecto *Excesos sin castigo o Memoria robada*? Si su respuesta es negativa, marque y pase a la pregunta número 15.

- () Sí
 () No

14. ¿Cuántas bases de datos combinó?

- () 2
 () 3
 () 4
 () Otro: _____

Análisis de datos

15. ¿Participó en la fase de análisis de datos de la investigación *Excesos sin castigo* o *Memoria robada*? Si la respuesta es negativa, marque y pase a la pregunta número 17.

Sí

No

16. ¿Qué programas y/o herramientas empleó para el análisis de datos? Complete su nivel de habilidad: B (básico), I (intermedio) o A (avanzado). Puede marcar más de una opción.

Excel:

- Nivel B I A

Open Refine:

- Nivel B I A

R:

- Nivel B I A

Otro (especifique el nivel): _____

Verificación de datos

17. ¿Participó en la verificación de la información contenida en las bases de datos de la investigación *Excesos sin castigo* o *Memoria robada*?

Sí

No

Comunicación de datos

18. ¿Qué lenguajes de programación y/o herramientas empleó para el diseño y/o implementación del sitio web? Indique su nivel de habilidad: B (básico), I (intermedio) o A (avanzado). Puede marcar más de una opción.

HTML:

- Nivel B I A

Java:

- Nivel B I A

Python:

- Nivel B I A

Indesign:

- Nivel B I A

Photoshop:

- Nivel B I A

Illustrator:

- Nivel B I A

No participé en este proceso

Otro (especifique el nivel): _____

19. ¿Qué lenguajes de programación y/o herramientas empleó para el diseño y/o implementación de visualizaciones? Indique su nivel de habilidad: B (básico), I (intermedio) o A (avanzado). Puede marcar más de una opción.

() HTML:

- Nivel B () I () A ()

() Java:

- Nivel B () I () A ()

() Python:

- Nivel B () I () A ()

() Indesign:

- Nivel B () I () A ()

() Photoshop:

- Nivel B () I () A ()

() Illustrator:

- Nivel B () I () A ()

() No participé en este proceso

() Otro (especifique el nivel): _____

¡Gracias por completar este cuestionario!

Anexo 3: Entrevista a la periodista Milagros Salazar (MS) de *Convoca*

Fecha: 7 de setiembre de 2019

Duración: 55 minutos

Entrevistadora: Rocío Romero (RR)

Recopilación de datos

RR: ¿La obtención de los documentos fue principalmente por la Ley de Transparencia y Acceso a la Información Pública?

MS: Hubo un componente de la Ley, sobre todo a lo largo del proyecto, pero había información que descargamos del OEFA porque tiene un buscador para poder acceder a los procesos sancionadores. Paralelamente hacíamos pedidos de información para profundizar en los expedientes. O sea, si había un caso que nos llamaba la atención, solicitábamos el expediente, fotografías, videos, los informes completos para poder documentar el caso.

RR: ¿Cuántos pedidos de acceso a la información realizaron?

MS: Fueron como 130.

RR: ¿Si en Perú no hubiera existido la Ley de Transparencia y Acceso a la Información Pública, el resultado de la investigación habría sido el mismo?

MS: La Ley de Transparencia y Acceso a la Información Pública fue muy importante porque, por ejemplo, nosotros llegamos a obtener los mil informes ambientales encarpados. Si bien es cierto demoró la entrega, debido a las barreras burocráticas o la jerga burocrática para traducir la información, fue fundamental. Tenemos una Ley que, con sus debilidades, permitió hacer el proyecto. Pero eso es porque somos un equipo que se dedica a hacer esto, pero no es viable para un periodista de un medio tradicional dedicar el tiempo a esto. La Ley cansa a los periodistas.

Limpieza de datos

RR: ¿Quiénes participaron en la estructuración de las matrices de datos?

MS: Una cosa era la matriz para la investigación que después se transformó en una matriz para la visualización. La matriz para la investigación la construí yo con Esteban Valle Riestra, porque es muy bueno trabajando con datos: viene de las Ciencias Políticas y él hacía operaciones más complejas que un periodista en promedio. Y ya para la visualización trabajamos con Víctor Anaya. Víctor se encargó de la estructura de la base de datos para la herramienta, para la aplicación web. Entonces, como que la matriz para la investigación se creó como base fundamental, pero son dos miradas diferentes.

RR: ¿La otra matriz de datos era más pequeña?

MS: Más pequeña, menos técnica, pero es por el entorno. Nosotros hicimos mapas y ahí había que ubicar las coordenadas, georreferencias y ese trabajo lo hizo muy bien Aramís. Él tenía que entrar a los proyectos mineros, buscar las coordenadas de cada proyecto, ahí había otra fuente de información. Entonces es un trabajo de hormiguita.

RR: ¿Cómo fue la labor de los voluntarios para la transcripción de información?

MS: Creo que hice una convocatoria abierta o fue a partir de unos talleres que dicté con Aramís. No me acuerdo muy bien. La cuestión es que nos fijamos que había muchos jóvenes, estudiantes míos y los convoqué. Ellos no cumplían con una jornada, ellos o trabajan de su casa o iban dos horas a llenar la base de datos. Era una actividad muy puntual: llenar su data, porque en el proceso nos dimos cuenta de que era inviable que dos o tres personas hicieran esa chamba. Fue un trabajo demasiado duro porque, claro, un joven con poca experiencia no llena una base de datos como debería llenarla a veces, por más capacitación que le des. Entonces, nos tomó mucho tiempo la limpieza, sobre todo a Aramís. Después de Aramís, yo tenía que limpiar, había más filtros y más filtros y una base de datos sucia se convertía en un dolor de cabeza tremendo.

Entonces, ahí también nosotros aprendimos del proceso: más gente en un equipo no te garantiza calidad del trabajo. Y lo otro es que tienes que tener muy claro los procesos y en todo proceso tener un control de la calidad de los datos porque luego se vuelve un dolor de cabeza. Pero nos sirvió a nosotros y de hecho también sirvió para que una nueva generación de jóvenes se pudiera familiarizar con una forma de periodismo, era como una escuela también para ellos.

Verificación de datos

RR: ¿Hubo un proceso de verificación de la información obtenida?

MS: Sí.

RR: ¿Ese proceso fue parcial o total?

En principio, estás trabajando con información oficial: nombres, registros de procesos de sanción que el Estado registra. Cuando me refiero que hacíamos un chequeo total de verificación, me refiero a todo el proceso de trabajo. La verificación se hace antes, durante, después, todo el tiempo estás verificando. Entonces, yo te diría que todo aquello que se publicó estaba verificado y que en el proceso hubo una serie de verificaciones. Por ejemplo, si teníamos el ranking de las más recurrentes, se llamó a todas las empresas para chequear esa información o al tener los casos segmentados, chequear al detalle sus casos. Aquello que fue publicado estaba chequeado.

Comunicación de datos

RR: ¿Ustedes planificaron una estrategia de difusión *online* y *offline*?

MS: Mira, lo que hicimos fue alianzas con el diario *La República* para publicar e hicimos un trabajo en redes sociales. Tuvimos un tropiezo en el proceso. Paralelamente al proyecto, estábamos elaborando nuestro sitio web, el sitio web de *Convoca*, y lamentablemente la persona que contratamos para el servicio no hizo un buen trabajo.

Entonces, la página era lenta, tenía problemas de estructura, habíamos planificado una hora de lanzamiento y tuvimos problemas en la web. La planificación que tuvimos no se pudo cumplir justamente por el problema de origen de la página. Recuerdo que empezamos a publicar en una hora no muy amigable.

RR: ¿Ya estaba coordinado con *La República*?

MS: No. Es que el primer reportaje no salió con *La República*, sino los siguientes. El primer reportaje salimos solo nosotros. Incluso *IDL-Reporteros* lo difundió también porque ahí estaba Gustavo Gorriti también asesorando. Entonces, y claro lo lanzamos, creo que fue un martes o miércoles, en la tarde, porque habíamos planificado salir 8 de la mañana y no levantaba, había problemas. Eran las 12, yo me tenía que ir a dictar clases y los chicos lo querían lanzar. Yo dije lo lanzaremos desde la Universidad Ruiz de Montoya y así lo hicimos. En ese momento no sabíamos la gravedad de la página, pensábamos que era algo temporal, pensábamos que después iba a correr mejor. Fue todo un reto y nos salvó bastante el hecho de que Víctor hiciera en un *site* especial la herramienta que no dependía de la página principal, sino colapsaba porque había varias búsquedas. Nos trajo mucho aprendizaje porque confiamos en esta persona que había creado un medio y había sido recomendado por un miembro de *Convoca*. Confiamos bastante, pero bueno así pasa con las recomendaciones. Entonces, sí había una estrategia, pero tuvimos estos asuntos tecnológicos que impidieron salir a la hora planificada y luego la estrategia básicamente fue redes, emails, alianza con un medio impreso porque queríamos llegar al interior del país. Como era una investigación sobre regiones, nos parecía que esa era la manera de llegar.

Otras preguntas

RR: ¿Cómo surgió la idea del proyecto Excesos sin castigo?

MS: La idea surgió en julio de 2014 con la aprobación de una norma en el gobierno de Ollanta Humala que fue conocida como el paquetazo ambiental, la Ley N° 30230. Para

alentar la inversión en minería, petróleo e industrias extractivas se optó por dar una especie de amnistía ambiental. Si había un proceso sancionador en el organismo regulador, que hasta ahora sigue siendo el OEFA que depende del Ministerio del Ambiente, en la primera instancia las multas por infringir las leyes ambientales quedaban suspendidas. Era cero multas, pero sí se debía implementar medidas correctivas y, si estaba en una segunda instancia el proceso sancionador, se reducía el pago de la multa al 50 % por infringir las normas.

Hubo excepciones. En casos en los que la infracción atenta contra la vida o afecta la salud de las personas, eso no procedía. Pero se prestaba a una interpretación muy amplia para saber en qué casos no aplicaba. De la revisión que luego hicimos, casi en todos los casos se reducía la multa. Se suponía que esto era una medida temporal solo por tres años.

Entonces, nosotros nos hicimos varias preguntas. Primero si esto era resultado de lobby, si había un sector de las empresas mineras y petroleras que resultaba beneficiado, si ellas habían propuesto esta iniciativa al gobierno de Ollanta Humala y de qué manera se había influenciado en la toma de decisiones. No encontramos suficientes elementos para probar eso. Entonces, lo que decidimos hacer fue documentar el impacto de esta decisión pública, de esta decisión de Estado, calculando a cuánto iba a equivaler este monto de dinero que no se iba a cobrar en multas a estas empresas.

Hicimos un análisis de 8 a 9 meses a partir de las resoluciones de sanción del propio OEFA que decían esta empresa ha infringido tal número de normas, lo cual equivalía a cobrar tanto en UIT (Unidades Impositivas Tributarias), pero debido a la aprobación del artículo tal de la Ley N° 30230, esta multa no será cobrada. Entonces, este cálculo era hecho por el propio OEFA, donde renunciaba a su función fiscalizadora por esta norma aprobada en el gobierno de Humala, impulsada desde el Ejecutivo y que finalmente fue aprobada en el Congreso.

Ese fue el primer acercamiento, la razón de hacernos preguntas más allá de la aprobación de la norma, porque había de hecho una historia de la opinión pública, sobre todo de las

organizaciones que trabajan en la defensa de los derechos ambientales o que acompañan a las poblaciones afectadas por la minería. Pero nosotros decíamos el periodismo no se basa en la especulación ni en las teorías de las conspiraciones. Tenemos que hacer que la información hable por sí sola y demuestre si hay lobby, no hay lobby, cuál es el impacto, cuáles son los casos, cuáles son las empresas, de qué manera impacta esto a las poblaciones, que son las preguntas de las que parte un periodista.

En el proceso también había otro aspecto importante: los miembros de *Convoca*, sobre todo yo, veníamos investigando industrias extractivas hace varios años atrás. El embrión de *Excesos sin castigo* es una investigación que publiqué el 2010 en *IDL-Reporteros* y que se llama *La impotente regulación*.

Yo hice una revisión de varios años de los procesos sancionadores de las empresas mineras. Yo no trabajaba construyendo bases de datos, era una cuestión más de reportería tradicional. Entonces, yo ya tenía el antecedente de conocer quién es quién en las industrias extractivas y cómo se comportaba el sector a nivel de regulación. Tenía ese *background* y entonces por un lado como que el contexto era la norma, pero a la vez había capacidades dentro del equipo para poder mirar el sector que es muy especializado y técnico. Y el tercer aspecto era que es la primera investigación de *Convoca* con la que planeamos salir a la luz. El plan era que esa fuera la primera, pero realmente fue *Swiss Leaks*, una alianza con el Consorcio Internacional de Periodistas de Investigación y luego publicamos esta. Pero sí fue la primera del equipo.

Yo estaba como muy convencida, a partir de una experiencia que había tenido investigando otro sector económico que es la pesca, que la metodología que habíamos decidido utilizar para esta investigación era precisa para poder mostrar cómo funciona la fiscalización ambiental, los intereses que puede haber ahí, y el impacto que hay en una débil fiscalización.

Entonces, era la conjunción de tres aspectos: la coyuntura de ese momento buscando hacerse preguntas distintas que respondan qué hay detrás de esa norma y cómo va a impactar a la sociedad, al país; la segunda era el conocimiento que había del sector: de las industrias extractivas y básicamente la fiscalización ambiental por las capacidades del equipo; y la tercera era la voluntad de hacer una serie investigativa que no fuera una cobertura solo de un caso o cotidiano, sino que mostrara algo que normalmente en el periodismo peruano no se solía hacer. Yo diría que así nació el proyecto.

RR: ¿Este proyecto fue financiado?

MS: Sí. Nosotros obtuvimos fondos de *Oxfam*. Conseguimos un financiamiento aproximado por ocho meses.

RR: ¿Era efectivo?

MS: No, eran recursos. Nosotros en ese momento todavía no habíamos fundado, creado, constituido jurídicamente *Convoca*. *IDL* fue la organización que recibió los recursos y administró nuestros fondos, pero sí lo obtuvimos nosotros.

RR: ¿De cuánto era el fondo?

MS: El monto total era de 50 mil dólares, pero *IDL* cobraba un porcentaje, el 15 % más o menos por administrar. Al final era 35 mil dólares. Eso nos sirvió para pagar eso y las cosas que también hacíamos en *Convoca*. Era como un fondo semilla para el nacimiento de la organización. Gran parte se usó para los programadores. Pero el tema fue tan grande que se prolongó y después había que buscar otros fondos.

Anexo 4: Entrevista a la periodista Fabiola Torres (FT) de *Ojo Público*

Fecha: 6 de setiembre de 2019

Duración: 55:44 minutos

Entrevistadora: Rocío Romero (RR)

Recopilación de datos

RR: ¿El proyecto hubiera tenido el mismo resultado si en Perú no hubiera existido la Ley de Transparencia y Acceso a la Información Pública?

FT: Hubiera sido muy difícil como lo fue en México. Por ejemplo, en México Tania Montalvo no hizo solicitudes. Tuvo como respuestas que esa información era secreta o estaba embargada hasta un número de años. Eso, por ejemplo, impidió que México construyera una base de datos como la que nosotros hicimos y eso que también tenía este problema de robo de bienes de una manera masiva. Lo sabíamos porque muchos reportes anteriores a *Memoria robada*, en formato de reportajes, mencionaban los robos en decenas de iglesias de México, pero no mencionaban cuántas de estas piezas habían estado registradas, cuántas denunciadas, nada de eso. La precisión no existía. Solamente se basaba en lo declarativo de las autoridades responsables en ese momento. Eso fue bien difícil.

RR: ¿Cuántas solicitudes de acceso a la información pública se hicieron?

FT: Cerca de 117. Además, se hicieron reconsultas a las solicitudes. En ningún otro proyecto habíamos hecho tantas porque aparentemente la información existía, pero no sabíamos en qué términos preguntarla.

Limpieza de datos

RR: ¿Quién estuvo a cargo de la estructuración de las bases de datos?

FT: Nosotros nos hemos reunido siempre para ver bases de datos entre José Luis Huacles, tú y yo. Y luego hemos hablado con Jason cuando empezamos a diseñar el proyecto y qué se

quería visualizar. Era muy instintiva nuestra organización de los datos. Instintiva porque no es que habíamos partido de una ficha y cierta metodología. Eso lo fuimos afinando en el camino o incluso nos dábamos cuenta de que arrastrábamos cosas que no se habían llenado, que no se habían considerado en el proceso mismo. Lo que sí teníamos muy claro era la necesidad de organizar los datos, de pasar los documentos a Excel. Entonces José Luis en muchos de los casos se encargaba de crear las estructuras y luego validarlas contigo. Luego preguntarle a Johana que también llenaba bases de datos e ir combinando las tareas de recolección de información. Después cuando consiguió Audrey Córdova la base de datos a la Interpol ahí pedimos la ayuda de Jason porque creo que *scrapeo* algunas cosas del Ministerio de Cultura. Fueron muy puntuales los *scrapeos*. Más hubo una lectura de documentos enorme.

RR: ¿Por eso la necesidad de tener voluntarios?

FT: Sí, porque las fichas necesitaban imágenes. Yo recuerdo perfectamente que una de nuestras reuniones, como en los últimos dos o tres meses antes de lanzar el proyecto, teníamos que delimitar una base de datos porque no íbamos a acabar nunca, porque las solicitudes venían respondiéndose. Luego vimos que la necesidad de las imágenes en las fichas, en lo que habíamos construido, era súper importante. Entonces vinieron dos voluntarios para poder hacer toda la base de datos e imágenes que bajo tu supervisión porque ellos hicieron sus prácticas con ese trabajo. Ellos trabajaron en la data al comienzo, luego vimos que no teníamos garantías de que fueran a llenar correctamente y se les encargó a los chicos las imágenes, encargarse de construimos el registro de imágenes que les pedíamos que capturaran, que buscaran. Era un ejercicio muy manual. Creo que una cosa que ha caracterizado a ese proyecto es que hemos actuado de una manera bien instintiva y teníamos todo a nuestro favor porque no existe data y crearla ya es un gran avance. Es un

triunfo. Contestaron muchísimas solicitudes y en los pedidos denegados también había muchas respuestas.

Verificación de datos

RR: ¿Hubo un proceso de verificación de datos?

FT: Yo creo que el proceso mismo de leer los documentos y de entender qué estaba ocurriendo con esas piezas que se estaban reclamando o dadas por robadas tenía en realidad casi un proceso de descubrimiento. De “ay, caramba, existía esta pieza” y no está en los museos de Perú o fue saqueada de tal iglesia. No hicimos un ejercicio de verificación de cada pieza, creo que hubiera sido bastante complicado porque eran bastantes, pero creo que lo que sí ocurrió para casos muy puntuales. Por ejemplo, los bienes que llegaron a la exposición de bienes repatriados. Ahí por ejemplo supe bien la historia de qué bienes se trataban, qué culturas preincaicas se trataba, de la Cultura Lima, eso no lo teníamos muy claro a partir de los datos que teníamos. Incluso ese reporte creo que Cultura lo tenía bien general hasta que llegó e hicieron incluso una exposición pública, pero nosotros lo reconstruimos y ahí sí vino un proceso de verificación porque pudimos leer los documentos de Argentina y ahí figuraban detalles que había omitido, que no necesariamente los tenían en cuenta al momento de hacer los reportes de qué se había repatriado.

Comunicación de datos

RR: ¿En el proyecto hubo una estrategia de difusión *online* y *offline*?

FT: Lo conversamos, pero creo que las últimas semanas fueron muy intensas que no permitieron sacar adelante como se debía. Recuerdo que la siguiente semana del lanzamiento tenía que viajar a México. Y ahí expuse el proyecto. Lo que sí sugerí fue enviar una nota de prensa a todas las organizaciones periodísticas. No sé si se llegó a enviar, la redactamos me acuerdo. Y además hicimos el cómo se hizo que no se publicó. Pero esa

parte no se ejecutó con todo lo que se pensó, igual que toda la línea gráfica para difusión, porque se hizo muchísimo.

Otras preguntas

RR: ¿Cómo surgió la idea del proyecto *Memoria robada*?

FT: El tema surgió a partir de una conversación que tuvimos con David Hidalgo acerca del problema del tráfico de bienes culturales. A fines del 2015, él se encontraba haciendo una investigación para su libro. Hablamos de que el Ministerio de Cultura no tenía organizada todas las cifras, en principio hablábamos de cifras, la cantidad de robos. Conversamos la idea de empezar a recopilar todos los reportes de los robos, de las alertas sobre ventas de piezas arqueológicas en casas de subasta.

Lo que teníamos en ese momento era una idea de que el Ministerio de Cultura tenía reportes dispersos y no teníamos el universo de organizaciones o entidades del Estado que podían ser responsables. Entonces, empezamos a hacer una recopilación de información para crear un proyecto de datos.

La idea del proyecto de datos la escribimos con David Hidalgo para postular al fondo de la *International Women's Media Foundation*. Ahí creamos un proyecto con una mirada regional, que sea transnacional, porque una de las cosas que era evidente era que el tráfico de bienes culturales cruzaba fronteras y nosotros como país, que tiene una riqueza cultural muy grande, con muchas piezas, estábamos siendo afectados por no tener una base de datos.

También habíamos leído que las fichas eran importantes en tanto se tuviera un registro fotográfico. La ambición era un censo primero local de bienes culturales robados, subastados, repatriados, con la idea de que luego se convirtiera en un censo latinoamericano.

Por eso animé a tres periodistas que yo conocía bien: a Catalina Oquendo, quien había investigado en Colombia el tema del tráfico de bienes culturales; a Tania Montalvo, quien había hecho *Narcodata* y trabajaba con bases de datos; y a Alejandra Gutiérrez de Plaza

Pública. Y David invitó a Hassel Fallas que también hacía periodismo de datos para conformar el equipo.

La decisión de hacer el proyecto como tal se dio cuando confirmaron el fondo que fue en marzo o abril de 2016. Veníamos pidiendo solicitudes de acceso a la información a fines de 2015 y comienzos de 2016 porque ya existía la idea incipiente de que sí se tenía que hacer una base de datos. Así todo va evolucionando con las limitaciones que hubo para ejecutar el proyecto, así como las nuevas ideas que surgieron como; por ejemplo, hacer un plan de trabajo en el que no necesariamente los resultados iban a basarse en el análisis de datos.

A mí me tocó hacer un ángulo que no se basaba en la base de datos. Era sobre los traficantes, pero yo creé un mini registro, una lista básicamente, de casos antiguos y me puse a hacer una especie de ejercicio de arqueología periodística de mirar en qué habían quedado esos casos. Entonces, ahí encontré un expediente, que fue el caso que pude reportear con Catalina Oquendo sobre un anticuario que está en Buenos Aires, en San Telmo, y fue a la Fiscalía y todo. Pero ahí el ejercicio fue hacer una lista.

Si nosotros teníamos un universo de datos, este era una unidad de los datos, un listado de casos de tráfico que surgió para poder delimitar este tema, pero me sirvió mucho luego poder ver la dimensión del problema con la cantidad de subastas que había. Aquí el dato más relevante que me dio una idea de la dimensión que no estábamos viendo era cuántas casas de subastas estaban ofreciendo piezas culturales peruanas. Nosotros solo sabíamos de *Sotheby's* y *Christie's* porque lo pasaban en las noticias, pero no teníamos idea de que había 19 o 21 identificadas y que salieron en los reportes. Para mí, esa fue una línea de investigación que aún no termino porque hubo una historia que no se publicó, pero si hay algo súper valioso del proyecto a nivel de abrir esos datos y reconstruirlos, fue la dimensión del delito porque al final se estaba dando legalidad a piezas robadas en casas de subastas muy conocidas y prestigiosas.

RR: ¿De cuánto era el monto del fondo?

FT: 19 mil 900 dólares. Con los impuestos un poquito menos.

RR: El proyecto comenzó a mapearse a fines del 2015, ¿en qué mes inició?

FT: Formalmente el 2016 porque con David Hidalgo teníamos claro que ese año íbamos a presentar el proyecto a un fondo. Fue una buena decisión porque además de este fondo que gané, que fue el que ayudó a sostener financieramente ese proyecto, él también presentó una solicitud de para un grant más pequeño, como de 5 mil dólares a la *Fund for Investigative Journalism*. Entonces esa fue una decisión de apostarle a este proyecto y por eso también lo diseñamos así y caminó a inicios de abril financieramente. Por esas fechas conformamos un equipo transnacional.

Anexo 5: Entrevista al periodista David Hidalgo (DH) de Ojo Público

Fecha: 3 de junio de 2019

Duración: 41:16 minutos

Entrevistadora: Rocío Romero (RR)

Recopilación de datos

RR: ¿El proyecto se hubiera desarrollado de la misma manera si en Perú no existiera la Ley de Transparencia y Acceso a la Información Pública?

DH: No, porque hubo un gran trabajo de pedidos de información. Primero entendimos cómo funcionaba el esquema, empezamos a hacer pedidos de información de manera aleatoria al inicio y luego el aprendizaje más interesante de este proyecto fue que al leer los documentos uno podía entender la lógica del sistema burocrático del Estado. Por ejemplo, plantear informes que están relacionados a otro tipo de informes, a actas, a una serie de documentos que tienen diferentes instituciones públicas, de diferentes estamentos. Entonces, uno entendía cómo funcionaba esto, podía seguir la ruta en que el Estado había detectado, investigando o, simplemente, abandonado casos emblemáticos. Eso solamente fue posible porque empezamos a descubrir maneras de aprovechar esa Ley que además tiene algunos vacíos, pero que sí permitía una base sobre la cual trabajar. Ir pidiendo diversos tipos de información que además estaba mencionada de manera muy tangencial en la propia web del Ministerio de Cultura; es decir, resúmenes con datos muy aislados y lo que nosotros hicimos fue ver solicitarlos de manera integral para poder hacer nuestro propio análisis. Y una vez que empezamos a recibir los datos también hicimos pedidos de los documentos que sustentaban esos datos y al analizarlos vimos que estaban relacionados a otros documentos y se pudo hacer una lectura general.

RR: ¿Cuál fue el principal reto en esta fase de recopilación de datos aplicando la Ley?

DH: Había que saber qué tipo de información había en el Estado y por eso empezamos con una etapa de solicitudes de información bastante intensa en donde estuvo todo un equipo y luego se empezó a hacer el análisis. Fuimos armando una estrategia para superar los baches que el mismo Estado nos ponía. Por ejemplo, cuando uno pide los informes de robos en determinadas instituciones pueden responder no existen informes de robos. ¿Por qué? Porque esa institución no les llama informes, sino reportes. Había que cambiar simplemente la expresión para poder obtener información correcta. Esas pequeñas triquiñuelas o zancadillas que pone el Estado para entregar la información tuvieron que ser analizadas y respondidas con una estrategia periodística que apuntaba a llenar los vacíos de preguntas que nos habíamos hecho en una segunda etapa.

RR: ¿Qué importancia tuvo el recurso de hábeas data para el tema periodístico?

DH: Es fundamental, pero lamentablemente en nuestra experiencia no fue buena porque nunca recibimos esa información. Era esencial porque nosotros teníamos la información de que hay un procedimiento por el cual se emiten documentos que permiten blanquear el origen de ciertos objetos o disfrazarlos, que es lo mismo que ocurre en el tema de tráfico de madera, se emiten documentos oficiales, con sellos oficiales para limpiar el origen de madera ilícita y otro tipo de actividades comerciales que también utilizan información falsificada, adulterada en base a formatos oficiales para alimentar una actividad ilícita. En este caso teníamos indicios y simplemente el Estado se amparó en que eso pertenecía al ámbito del secreto comercial; sin embargo, con abogados que consultamos en la época no correspondía a la realidad, y a pesar de que planteamos este recurso no lo obtuvimos. Si lo hubiéramos obtenido, creo que hubiéramos tenido una nueva veta de investigación sobre personas involucradas porque una de las tareas y metas de este proyecto era identificar responsables que es, digamos, el sentido del periodismo de investigación.

Limpieza de datos

RR: ¿Qué tan importante fue para el proyecto periodístico incluir a un programador en la estructuración de las matrices de datos?

DH: En el diálogo con este programador aprendimos una manera más efectiva de estructurar los datos, que pudiera ayudar a visualizar más rápidamente o más claramente la información o de manera más efectiva, eliminando datos que no eran necesarios.

Verificación de datos

RR: ¿Hubo un proceso de verificación de datos?

DH: Hubo un proceso; por ejemplo, para cotejar los datos de los informes de subastas en el extranjero. Entonces, hubo una revisión de las páginas webs originales de las casas de subastas donde se mostraban estos objetos. También hubo un trabajo para buscar documentación en estudios históricos y arqueológicos sobre cierto tipo de materiales que estaban siendo subastados en el extranjero. Es decir, para ver los estilos por periodos y a nivel de los datos mismos pues hubo un trabajo de limpieza y de sí, hasta cierto punto, cotejo porque en algunos casos las bases de datos que te ofrecía el Estado no necesariamente estaban limpias, había que revisar si la documentación en papel sustentaba lo que decían los datos. Entonces, sí hay un trabajo de cotejo y de verificación.