



DATAista

Sistema de Información para la incursión al Periodismo de Datos



Impreso en México

CDMX, 28 de noviembre de 2017.

**Dataísta: Sistema de Información
para la Incursión al Periodismo de Datos**

UAM Cuajimalpa, MADIC

Dataísta: Sistema de Información para la Incurción al Periodismo de Datos

Presentan:

Alazañez Cortés Belén
Miranda Aguirre Zayra Monserrat
Molina Barradas Jocelyn Lizbeth
Monroy Cuevas Erick

Asesores:

Dra. Rocío Abascal Mena
Dr. Román Esqueda Atayde
Dr. Rodrigo Gómez García

Lector interno:

Dr. César Augusto Rodríguez Cano

Lector externo:

Dr. Irving Omar Morales Agiss



Universidad Autónoma
Metropolitana



División
de Ciencias
de la Comunicación
y Diseño

MADIC

Maestría en Diseño,
Información y Comunicación

Agradecimientos

“De manera conjunta, agradecemos a nuestros asesores, la Dra. Rocío Abascal, el Dr. Rodrigo Gómez y el Dr. Román Esqueda por su paciencia, consejos y guía durante este proceso de investigación; asimismo agradecemos a nuestros lectores el Dr. Irving Morales y el Dr. César Augusto Cano por su tiempo y aportaciones al proyecto. A todos los periodistas que colaboraron con nosotros en cada una de las fases de la investigación y a SocialTIC, sin ustedes “Dataísta” no habría sido posible. A la Universidad Autónoma Metropolitana, Unidad Cuajimalpa y a la MADIC por proporcionarnos los conocimientos necesarios para poder concluir satisfactoriamente otra etapa de nuestra vida académica.”

Belén, Erick, Jocelyn y Zayra

A mi madre por enseñarme que con trabajo duro y pasión se pueden cumplir todas las metas.
¡Gracias por creer en mí!

A mi padre por su guía y apoyo durante todo este proceso.

A Roberto por las risas, por impulsarme y mostrarme el camino.

A David por las palabras de aliento después de un día complicado.

A mis tías Orquídea y Antonieta, por ser ejemplo en mi vida.

A Victor, por la paciencia y los consejos. ¡Gracias por estar!

A Belén y Arturo, sin ustedes la MADIC no habría sido lo mismo, ¡gracias por los momentos compartidos y las pláticas catárticas!

A mi equipo por el esfuerzo y la dedicación en nuestro proyecto.

Zayra Miranda

A mi madre por su apoyo, siempre incondicional.

A mi padre por legarme el interés por la educación y a Alejandro por hacerme más sensible a ella, reanimar mi curiosidad y multiplicarla.

A Zayra, por su entrega y transparencia en los momentos compartidos.

A mi equipo, por su dedicación al proyecto.

Belén Alazañez

A mis padres.

A Belén, Jocelyn y Zayra por su arduo trabajo y paciencia.

A la Dra. Rocío Abascal por su profesionalismo y dedicación a la MADIC.

Erick Monroy

Agradezco a mi esposo Uriel por todo su apoyo y amor.

A mi mamá por apoyarme siempre en cada decisión de mi vida.

A mis hermanos y mis sobrinos por siempre estar conmigo y su cariño.

A Belén, Zayra y Erick por su paciencia y esfuerzo en Dataísta.

A Lilia Saúl, Yosune Chamizo, Israel Piña, Irving Morales, Sergio Araiza y Gibran Mena por su tiempo, confianza y apoyo con nuestro proyecto.

Gracias papá por tu ejemplo y por haber sido un gran periodista.

Jocelyn Molina



Contenido

Introducción.....	9
Planteamiento del problema.....	10
Justificación.....	11
Objetivos.....	13
Objetivo General.....	13
Objetivos Específicos.....	13
Preguntas de Investigación.....	14
Hipótesis.....	15
Capítulo 1. Periodismo de Datos.....	16
1.1 Breve Historia del Periodismo de Datos.....	17
1.2 ¿Qué es el Periodismo de Datos?.....	22
1.3 Metodología del Periodismo de Datos.....	23
1.4 Impacto e intervención social de los nuevos medios en el periodismo.....	25
1.4.1 Periodismo y Hacking Cívico.....	29
1.5 Casos Internacionales de Periodismo de Datos.....	32
1.5.1 Chinaleaks.....	33
1.5.2 Panama Papers.....	34
1.6 Resumen del capítulo 1.....	35
Capítulo 2. Periodismo de Datos en México.....	37
2.1 Periodismo en México.....	38
2.2 Periodismo de Datos en México.....	41
2.3 Ejemplos de Periodismo de Datos en México.....	43

2.3.1 La Estafa Maestra.....	43	4.2.3 Big Data University.....	77
2.3.2 El (falso) éxito de la Cruzada contra el Hambre.....	44	4.2.4 DataCamp.....	78
2.3.3 Narcodata.....	46	4.2.5 DataBasic.IO.....	78
2.3.4 Desaparecidos.....	47	4.2.6 Tutellus.....	79
2.4 Resumen del capítulo 2.....	49	4.3 Resumen del capítulo 4.....	80
Capítulo 3. Datos Masivos y Datos Abiertos.....	50	Capítulo 5. Propuesta Interdisciplinaria.....	83
3.1 Datos Masivos.....	51	5.1 Interdisciplinarietà.....	84
3.2 Definiendo los Datos Abiertos.....	53	5.1.1 Estrategias en Comunicación.....	85
3.2.1 Clasificación.....	55	5.1.2 Diseño de Información.....	85
3.3 Datos Abiertos en México.....	56	5.1.3 Sistemas para la Interacción.....	86
3.3.1 El portal de Datos Abiertos de la República Mexicana.....	57	5.2 Metodología Aplicada en el Proyecto.....	86
3.4 Resumen del capítulo 3.....	59	5.2.1 Dimensión 1.....	86
Capítulo 4. Herramientas de apoyo para el Periodista de Datos.....	62	5.2.1.1 Hallazgos.....	87
4.1 Análisis de herramientas de apoyo para el Periodista de Datos.....	63	5.2.2 Dimensión 2.....	88
4.1.1 Herramientas de Búsqueda de datos.....	63	5.2.2.1 Hallazgos.....	89
4.1.2 Herramientas de <i>Scraping</i>	65	5.2.3 Dimensión 3.....	90
4.1.3 Herramientas de Limpieza de Datos.....	67	5.3 Usuario.....	91
4.1.4 Herramientas para la Exploración o Análisis de Datos.....	68	5.3.1 Perfil del Usuario.....	91
4.1.5 Herramientas para la Visualización de Datos.....	69	5.3.2 Análisis de Necesidades.....	91
4.1.6 Herramientas más usadas por los expertos.....	70	5.4 Resumen del capítulo 5.....	92
4.2 Plataformas Digitales para el Aprendizaje.....	72	Capítulo 6. Sistema de Información para la Incursión y el Aprendizaje de la Metodología del Periodismo de Datos.....	94
4.2.1 General Assembly.....	76	6.1 Requerimientos.....	95
4.2.2 Data Tells a Story: Reading Data in the Social Sciences and Humanities.....	76	6.2 Diseño Instruccional aplicado al Sistema de Información.....	95
		6.2.1 Dataísta desde la Perspectiva Constructivista.....	96
		6.3 Prototipo.....	97
		6.3.1 Mapa del Sistema.....	97

6.3.2 Modelado del Prototipo.....	99	6.8.6 Medios.....	121
6.3.2.1 Diagrama de Casos de Uso.....	99	6.8.7 Preguntas Frecuentes.....	121
6.3.2.2 Diagrama de Actividades	101	6.8.8 Foros.....	122
6.4 Dataísta.....	103	6.8.9 Inicio de Sesión o Registro.....	122
6.4.1 Concepto.....	103	6.8.10 Datos Abiertos.....	123
6.4.1.1 ¿Qué es?.....	103	6.8.11 Herramienta Auxiliar.....	123
6.4.1.2 Proposito del Sistema.....	103	6.8.12 Clasificador de Herramientas.....	124
6.4.1.3 Personalidad.....	103	6.9 Desarrollo de las Herramientas Auxiliares.....	124
6.4.2 Conceptualización del nombre.....	104	6.9.1 Python.....	127
6.4.3 Logotipo.....	104	6.9.2 Librería Flask.....	128
6.5 Diseño y arquitectura del sistema.....	106	6.9.3 Herramienta Obtención.....	130
6.5.1 Objetivos del sitio.....	106	6.9.4 Herramienta Limpieza.....	133
6.5.2 Bocetaje.....	106	6.9.5 Herramienta Análisis.....	133
6.5.3 Arquitectura del sitio.....	108	6.9.6 Herramienta Visualización.....	134
6.5.4 Diseño de la Interfaz.....	109	6.10 Resumen del capítulo 6.....	136
6.5.5 Color.....	110	Capítulo 7. Evaluación	138
6.5.6 Tipografía.....	111	7.1 Pruebas de Usabilidad.....	139
6.5.7 Botones, menú y elementos		7.2 Sesión de Focus Group.....	141
de la interfaz.....	111	7.3 Resumen del capítulo 7.....	150
6.6 Usabilidad.....	111	Capítulo 8. Conclusiones	152
6.7 Conceptualización y generación		Referencias.....	158
del contenido del sistema.....	112	Lista de ilustraciones.....	165
6.7.1 Selección del formato.....	112	Lista de tablas.....	168
6.7.2 Desarrollo de los contenidos.....	113	Anexo A: Tecnologías de la Información	
6.8 Descripción del Sistema de Información		en el Periodismo.....	169
Dataísta.....	114	Anexo B: Entrevistas.....	170
6.8.1 Descripción general del prototipo.....	117	Anexo C: Storyboard de la sección de Obtención.....	187
6.8.2 Página principal.....	117	Anexo D: Código Fuente Herramienta Obtención.....	190
6.8.3 Home.....	118	Anexo E: Código Fuente Herramienta Limpieza.....	195
6.8.4 Introducción.....	119	Anexo F: Código Fuente Herramienta Análisis.....	197
6.8.5 Secciones de Obtención,		Anexo G: Código Fuente Herramienta Visualización.....	199
Limpieza, Análisis y Visualización.....	119		



Introducción

En años recientes el periodismo ha enfrentado diversos retos y oportunidades, en buena medida, gracias a las innovaciones tecnológicas. La incorporación de las Tecnologías de Información y Comunicación (TIC) a nivel global, ha cambiado de forma significativa las distintas prácticas, rutinas y dinámicas relacionadas con el periodismo.

Hoy en día, el periodista se encuentra frente a una extensa gama de fuentes de información, entre ellas podemos encontrar bases de datos publicadas por los gobiernos, reportes de ONGs e incluso información recopilada de redes socio-digitales como Twitter, Facebook, entre otros.

Si bien, tener un abanico más amplio de fuentes de información es un gran apoyo para la investigación periodística, lo cierto es que, se requiere de una cierta especialización para poder procesar las grandes cantidades de información y datos que se pueden obtener y trabajar en los sistemas digitales. Ya no se trata de la metodología clásica del periodismo de investigación, donde el periodista reportaba y buscaba alguna documentación de gabinete para enriquecer su noticia o reportaje. Ahora, se pueden contar historias periodísticas sustentadas en datos cuantitativos, es decir, los datos toman un papel central en las investigaciones, no sólo fungiendo

como soporte, sino como una fuente fundamental. Esta mezcla entre lo tradicional y las posibilidades que permite la digitalización dio pie a la aparición del Periodismo de Datos.

La incorporación del Periodismo de Datos se ha dado de manera paulatina. Principalmente, porque los periodistas no cuentan con los conocimientos, el tiempo y la infraestructura necesarios para poder especializarse.

En ese sentido, el Periodismo de Datos, ha venido a dar un soplo de aire fresco a una industria sumergida en una profunda crisis, que se enfrenta a una generalizada y continua caída en circulación, lectura, audiencia, finanzas e incluso, credibilidad (Milan, 2015, pág. 16).

Por lo que el Periodismo de Datos es una oportunidad para recomponer o reinventarse la industria periodística. Sin embargo, es una labor que requiere de la integración de más periodistas ya que en la actualidad son muy pocos los medios en México que estén desarrollando Periodismo de Datos.

Por lo anterior, identificamos que es necesario dotar al periodista de herramientas que faciliten el aprendizaje de técnicas y programas específicos que permitan el entendimiento y aplicación de actividades propias del Periodismo de Datos.



Planteamiento del problema

Para contextualizar el tema del proyecto terminal es preciso responder a la siguiente pregunta;

¿Por qué es necesario un estudio en materia de Datos Abiertos y Periodismo de Datos en México?

Los Datos Abiertos son un fenómeno en crecimiento en Latinoamérica, el cual involucra principalmente a gobiernos y asociaciones de la sociedad civil. Sin embargo, paulatinamente se han estado incorporando nuevos actores como los periodistas, quienes en contextos de opacidad gubernamental y escasez de Datos Abiertos, han reunido repositorios de datos y desarrollado proyectos e iniciativas para compartir información públicamente (Casanueva, 2015).

Frecuentemente, el término Periodismo de Datos causa polémica, pues, tradicionalmente los periodistas han desempeñado su trabajo con ayuda de datos. Por ello, algunos especialistas sugieren denominarlo como Periodismo de Base de Datos (PBD), sin embargo, por razones de uso se prefiere utilizar el término más corto, es decir, Periodismo de Datos (P.D.) (Crucianelli, 2013).

El Periodismo de Datos es una disciplina emergente, en la cual los periodistas en conjunto con un equipo interdisciplinario formado por un analista de datos o programador y un diseñador combinan sus habilidades con el propósito de contar historias noticiables a través de información obtenida de grandes bases de datos abiertas, principalmente gubernamentales (Manual de Periodismo de Datos, 2012).

Lamentablemente, en México es complicado formar estos grupos interdisciplinarios principalmente por cuestiones financieras, pero también por la falta de un lenguaje común entre los miembros del equipo para trabajar en conjunto. Por ello es importante dotar al periodista de los conocimientos básicos para comenzar a implementar la metodología propuesta por el Periodismo de Datos en sus investigaciones.

Justificación

En 2011, el Centro Europeo de Periodismo (EJC), circuló una encuesta para conocer las necesidades de entrenamiento que tenían los periodistas sobre el Periodismo de Datos.

Dentro de sus hallazgos identificaron que había un creciente grupo de periodistas interesados en realizar investigaciones basadas en datos, desafortunadamente había una serie de vacíos tanto de conocimientos como de infraestructura que eran importantes de solucionar.

Es así que se requiere de una aproximación sistemática, pues usualmente se tiene una mala interpretación de números y estadísticas por parte de los periodistas. Extrapolar erróneamente las tendencias, malinterpretar desarrollos complejos y la falta de información son continuamente errores cometidos por los periodistas sin nociones sobre procesamiento de datos, por lo que es importante proveerlos de conocimientos y habilidades prácticas que les permitan realizar de forma certera investigaciones basadas en grandes cantidades de datos.

Este tipo de estudios no tiene algún símil en Latinoamérica o precisamente en México. Sin embargo,

en textos como el Manual de Periodismo de Datos (2012) se hace mención de la necesidad de los periodistas de especializarse; principalmente para ser capaces de aprovechar las nuevas fuentes de información que los avances tecnológicos están generando.

De acuerdo con Araiza (2015) son tres las deficiencias en el entorno de recursos sobre datos:

1. La mayoría de recursos no están en español.
2. Las referencias prácticas en el uso de los datos son referencias a sistemas diferentes al latinoamericano, es decir, en contextos tecnológicos, institucionales y de acceso a datos.
3. Existen recursos didácticos que se encuentran en varios espacios digitales y comunidades de práctica, pero es complicado acceder a recursos relevantes, requiriendo saber realizar ejercicios de búsqueda, selección y curación de información.

De acuerdo con Flores Vivar (2013), es de suma importancia para los periodistas de investigación que, antes de empezar a trabajar con aplicaciones concretas en el manejo de datos; entiendan de forma precisa las ventajas y desventajas del Periodismo de Datos, y que adquieran algunas habilidades generales para llevar a buen puerto los reportajes de investigación. En este sentido, nuestra propuesta llamada: Dataísta, se plantea como un Sistema de Información creado con la finalidad de proporcionar información, fácil y sencilla sobre la metodología y procesos del Periodismo de Datos. No pretende ser un tutorial o un curso masivo en línea; sino un espacio, en donde los interesados encuentren de manera rápida y eficaz sugerencias sobre cómo abordar un proyecto periodístico basado en grandes cantidades de datos digitales, es decir, que cuenten con un sistema básico que les permita entender el proceso del P.D.

Por lo tanto, la idea de Dataísta, es preparar al periodista de investigación en la aplicación de sus conocimientos a situaciones reales, nuevas y cambiantes; reforzando el aprendizaje significativo a partir del seguimiento de la metodología del P.D. El aprendizaje significativo se puede incorporar a las estructuras de conocimiento que tiene el sujeto y que adquiere significado a partir de la relación con conocimientos anteriores. Dataísta es un sistema, que surge a partir de las necesidades de los periodistas de investigación por aprender la metodología propuesta por el P.D. La metodología está compuesta por 4 etapas que son: obtención, limpieza o filtrado, análisis y visualización.



Objetivos

Objetivo General

Desarrollar un Sistema de Información que permita a los periodistas de investigación comprender el razonamiento propio de la metodología utilizada en el Periodismo de Datos para desarrollar periodismo de investigación sustentado en Datos Abiertos.

Objetivos Específicos

1. Identificar las principales barreras de los periodistas de investigación en la construcción de investigaciones periodísticas basadas en Datos Abiertos.
2. Aplicar la metodología del Periodismo de Datos en un sistema de información propuesto para el periodista de investigación.
3. Identificar la apropiación y la utilización de las bases de Datos Abiertos por parte de los periodistas de investigación para diseñar reportajes a través de la metodología del Periodismo de Datos.

Preguntas de Investigación

Con base en la información recabada, la pregunta central del proyecto es la siguiente:

¿Qué tipo de Sistema de Información puede ayudar a los periodistas de investigación a comprender la metodología y los procesos utilizados en el Periodismo de Datos?

A partir de ella, se derivan algunas sub-preguntas, las cuales representan una dimensión de la propuesta de investigación:

- ¿Cuáles son las principales dificultades que enfrentan los periodistas mexicanos de datos?
- ¿Qué metodología de trabajo deben seguir los periodistas mexicanos de datos?
- ¿Qué características debe tener el Sistema de Información para facilitar el aprendizaje de la metodología del Periodismo de Datos?



Hipótesis

La hipótesis propuesta para esta investigación es la siguiente:

La falta de conocimientos teóricos que permitan utilizar la metodología y procesos del Periodismo de Datos por parte de los periodistas de investigación mexicanos es la principal barrera para aprovechar las bases de Datos Abiertos liberadas por el gobierno. Por ello, un Sistema de Información permitirá la capacitación de los periodistas de investigación mexicanos en la aplicación de la metodología del Periodismo de Datos siendo el medio adecuado para facilitar la construcción de más investigaciones periodísticas basada en Datos Abiertos.

Periodismo de Datos

Capítulo **1**

Para poder entender el fenómeno que queremos abordar, así como las problemáticas que se desarrollan a su alrededor, es necesario explorar y revisar investigaciones sobre el P.D., los datos masivos y las herramientas digitales para periodistas.

Antes de abordar de lleno la historia del Periodismo de Datos, es importante tener claro el concepto de periodismo, disciplina que da origen a esta nueva especialización.

Existen diversas definiciones de lo que es el periodismo, sin embargo se decidió retomar aquellas que dejan clara su función social.

Alberto Dallal (1976) define el periodismo como *“el acto cabal de sociabilizar rápido y efectivamente la información”* (pág. 29).

Rivadeneira (1980) por su parte define al periodismo como *“un sistema abierto de comunicación humana tecnificada que procesa mediante acontecimientos, ideas y sentimientos procedentes de una o varias fuentes”* (pág. 30).

Por otro lado, el Instituto Americano de Prensa (s.f.) define al periodismo como *“la actividad de reunir, crear y presentar noticias e información. Es también el producto de estas actividades.”*

Finalmente, la Escuela Abierta de Periodismo (s.f.) dice que el periodismo es *“en su sentido más básico, el proceso de reunir y presentar información real. Esta información se utiliza para transmitir una historia al público y crear un registro con fines de archivo”*.

En síntesis, podemos decir que el periodismo es la actividad de recopilar, analizar e interpretar información relacionada con acontecimientos de interés social.

1.1 Breve Historia del Periodismo de Datos

Para un mejor entendimiento de esta nueva modalidad del periodismo es necesario conocer un poco sobre el desarrollo que esta actividad ha tenido a través del tiempo, para así poder situarlo en la actualidad.

Se dice que el periodismo siempre se ha desarrollado con base en datos, que es imposible imaginar una nota sin que esté basada en datos. Sin embargo, podríamos decir que la necesidad de que el periodismo vaya más allá de perseguir la noticia y recopilar declaraciones de funcionarios, surge en el primer tercio del siglo XX cuando diferentes periodistas presentaban la inquietud de hacer del periodismo, un oficio medible y comprobable. Max Weber, en Alemania, y Robert E. Park, en los Estados Unidos, plantearon el sentido complementario entre el trabajo del periodista y el científico social (Ávila, 2013).

Pasaron varios años para que se realizara el primer artículo periodístico desarrollado con base en metodologías científicas. Philip Meyer en 1959 fue quien se atrevió a dar ese primer gran salto, cuando cansado de que los periódicos estuviesen llenos de entrevistas a políticos y boletines de prensa oficiales, decidió crear una fuente propia, es decir, una base de datos. Philip Meyer, quien entonces trabajaba en *“The Miami Herald”*, eligió una queja ciudadana sobre los costosos que resultaban los seguros escolares para armar su base de datos. Tras cruzar diferentes

datos, descubrió que las compañías de seguros financiaban la mayor parte de la campaña presidencial estadounidense (Houston, 2016).

Después de la publicación de este artículo, Meyer decidió especializarse y gracias a una beca Nieman se dedicó a estudiar métodos de investigación social en la Universidad de Harvard (Ávila, 2013). Esto resulta relevante, porque en el 1967, luego de diversas marchas y disturbios que tuvieron lugar en la ciudad de Detroit, diferentes periódicos comenzaron a publicar que la mayoría de los manifestantes eran personas con baja escolaridad, que provenían de sectores marginales y grupos minoritarios. Meyer, quien para ese entonces ya trabajaba en *“The Detroit Free Press”* descubrió mediante la aplicación de encuestas que los manifestantes tenían estudios universitarios y eran de clase media. Gracias a estas dos investigaciones se considera a Meyer como el padre del periodismo de precisión o de investigación, precursor de lo que hoy se conoce como Periodismo de Datos o de bases de datos (Younge, 2011).

Años más tarde, en 1970 Philip Meyer publicó su libro *“Periodismo de Precisión”* en el que explica y defiende el uso de análisis de bases de datos y métodos de investigación social en el periodismo.

Durante los años subsecuentes diferentes periódicos comenzaron a desarrollar reportajes utilizando la metodología de Meyer, un ejemplo de ello es aquel publicado por *“The Washington Post”* en 1970 sobre las formas de reclutamiento de jóvenes durante la guerra de Vietnam.

Por su parte, en 1975 nació la organización *“Investigative Reports and Editors”* (IRE) cuya aportación al periodismo de investigación y de bases de datos ha sido fundamental para fomentar trabajos de fondo y con rigor verificable. Esta organización alcanzó notoriedad en 1976 cuando uno de sus fundadores, murió investigando temas de corrupción política y gansterismo. A partir de ese entonces la Universidad de Missouri abrió un espacio a la organización para garantizar su continuidad (Ávila, 2013).

A finales de 1980, alrededor de 50 periodistas en los Estados Unidos comenzaron a analizar datos para sus artículos, esto gracias a la aparición de las primeras computadoras personales y a un programa desarrollado por el periodista Elliot Jaspin y el programador – periodista Daniel Woods, llamado *“Nine Track Express”*. Este facilitaba la transferencia de cintas de computadora a las computadoras personales utilizando una unidad portátil (Houston, 2016). Es en ese momento cuando comienza a surgir lo que hoy conocemos como el Periodismo Asistido por Computadora.

Aprovechando el uso de las computadoras para el análisis de grandes bases de datos en 1989 el diario *“The Atlanta Journal”* publica diversos artículos sobre las disparidades raciales en los préstamos hipotecarios. Ese mismo año, Elliot Jaspin estableció en la Escuela Nacional de Missouri lo que hoy se conoce como el *“National Institute for Computer-Assisted Reporting”* (NICAR).

Como podemos advertir en los párrafos anteriores, todo el avance que estaba teniendo el periodismo se situaba en países anglosajones, principalmente en Estados Unidos. Fue hasta 1992, cuando las noticias sobre el periodismo de precisión llegaron al mundo de habla hispana con la publicación en “El País” (España) de un texto sobre el tema, bajo la autoría de Pedro Gómez y José Luis Dader. Al año siguiente este último tradujo al castellano y publicó el libro de Meyer “Periodismo de precisión, nuevas fronteras de la investigación periodística” (Ávila, 2013).

El retraso de los países latinoamericanos para utilizar las nuevas técnicas para el ejercicio del periodismo, se debe principalmente a que muchos de ellos se encontraban sumidos en una lucha por la libertad de expresión, pues muchos de sus gobiernos, principalmente dictaduras, impedían el ejercicio libre de la prensa (Ávila, 2013). Es por ello que para muchos de ellos sonaba imposible la especialización en bases de datos, los reportajes de fondo y la transparencia de estos.

La combinancia de la llegada del Internet así como el acceso a las fuentes de información, propiciaron un salto exponencial en el uso de la tecnología en el ejercicio del periodismo. Entre 2005 y 2007 comenzaron a surgir herramientas digitales y programas tanto gratuitos como de paga, para el manejo de bases de datos, mismos que ayudaron de manera significativa a que el ejercicio del Periodismo de Datos tuviera un mayor crecimiento, principalmente en países angloparlantes.

En 2009 surge el movimiento “*Hacks and Hackers*”, gracias al interés que tenían un grupo de periodistas en la convergencia entre periodismo y tecnología. Aron Pilhofer, de “*The New York Times*”; Rich Gordon, de “*Northwestern University*” y Burton Herman, ex corresponsal de AP, crearon una comunidad internacional para fomentar encuentros entre periodistas y expertos en tecnología para intercambiar ideas y encontrar colaboración entre ambas profesiones (Manual de Periodismo de Datos, 2012).

Por otro lado, el Centro Europeo de Periodismo, organizó en 2010 la primera conferencia sobre P.D., evento que tuvo lugar en Ámsterdam. Un año más tarde, en 2011, durante el *MozFest* celebrado en Londres, nace la iniciativa de escribir el primer Manual de Periodismo de Datos¹. El texto es una muestra del periodismo colaborativo, periodistas provenientes de diversos medios aportan sus herramientas y las comparten. En dicho manual se puede observar la participación de profesionales de *la BBC, del Chicago Tribune, de la Australian Broadcasting Corporation, La Nación (Argentina), The Washington Post, The Texas Tribune, Pro Pública, The New York Times, The Guardian*, entre otros.

Del 2000 a la fecha, diversas organizaciones internacionales han trabajado de manera conjunta para fomentar la realización de trabajos periodísticos basados en datos. Sin embargo, aún queda un largo camino por recorrer, sobre todo en los países de habla hispana.

¹ <http://interactivos.lanacion.com.ar/manual-data/index.html> (Consultado el: 3 de octubre de 2017)

A continuación se presenta una línea de tiempo que recapitula la historia del Periodismo de Datos:



Primer tercio del Siglo XX

Max Webner en Alemania y Robert E. Park en E.U. plantearon el sentido complementario entre el trabajo del periodista y el científico social.

1959

Philip Meyer realiza el primer artículo periodístico con base en metodologías científicas.

1970

Meyer publica su libro “Periodismo de precisión” y en adelante diferentes periódicos comienzan a utilizar su metodología.

1992

El periodismo de precisión llegó al mundo de habla hispana con la publicación en “El País” (España) de un texto sobre el tema.

Al año siguiente José Luis Dader tradujo al castellano y publicó el libro de Meyer “Periodismo de precisión, nuevas fronteras de la investigación periodística”.

2005 -2007

Comenzaron a surgir herramientas digitales y programas gratuitos y de paga para el manejo de bases de datos.

2009

Surge el movimiento “Hack and Hackers”, comunidad internacional para fomentar encuentros entre periodistas y expertos en tecnología para intercambiar ideas y encontrar colaboración entre ambas profesiones.

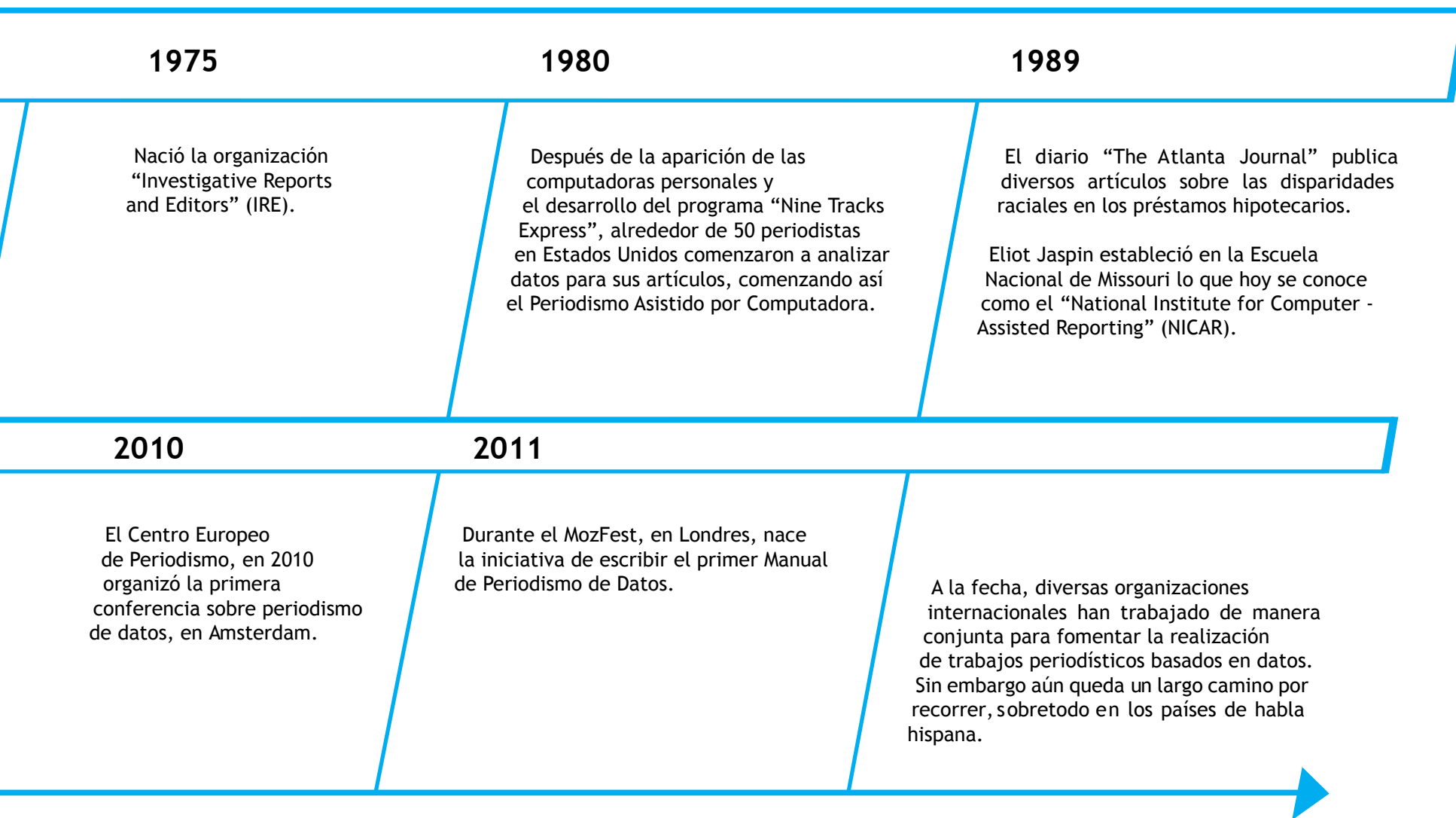


Figura 1. Línea del tiempo de la historia del Periodismo de Datos. (Fuente: Elaboración propia)

1.2 ¿Qué es el Periodismo de Datos?

En años recientes, cada vez es más frecuente escuchar las palabras “Periodismo de Datos”. De igual manera, existen más medios, sobre todo internacionales que dicen tener un área dentro de su redacción para el ejercicio del mismo, pero pocas veces se entiende a qué se refieren. A continuación, plasmamos una breve revisión de algunas de las definiciones que le dan al Periodismo

de Datos, diversos autores y/o periodistas.

Para Paul Bradshaw (2011), periodista independiente y académico de la Universidad de Birmingham; “el Periodismo de Datos es aquel en el que se utiliza el poder de las computadoras para encontrar, contrarrestar, visualizar y combinar información proveniente de diferentes fuentes”.

Por otro lado, el periodista español, Ignacio Ramonet ve al Periodismo de Datos como “otro tipo de periodismo, que está prosperando entre los nuevos nichos de Internet. Se le considera

Se presenta un diagrama donde se muestran los diferentes elementos que conforman el P.D.:

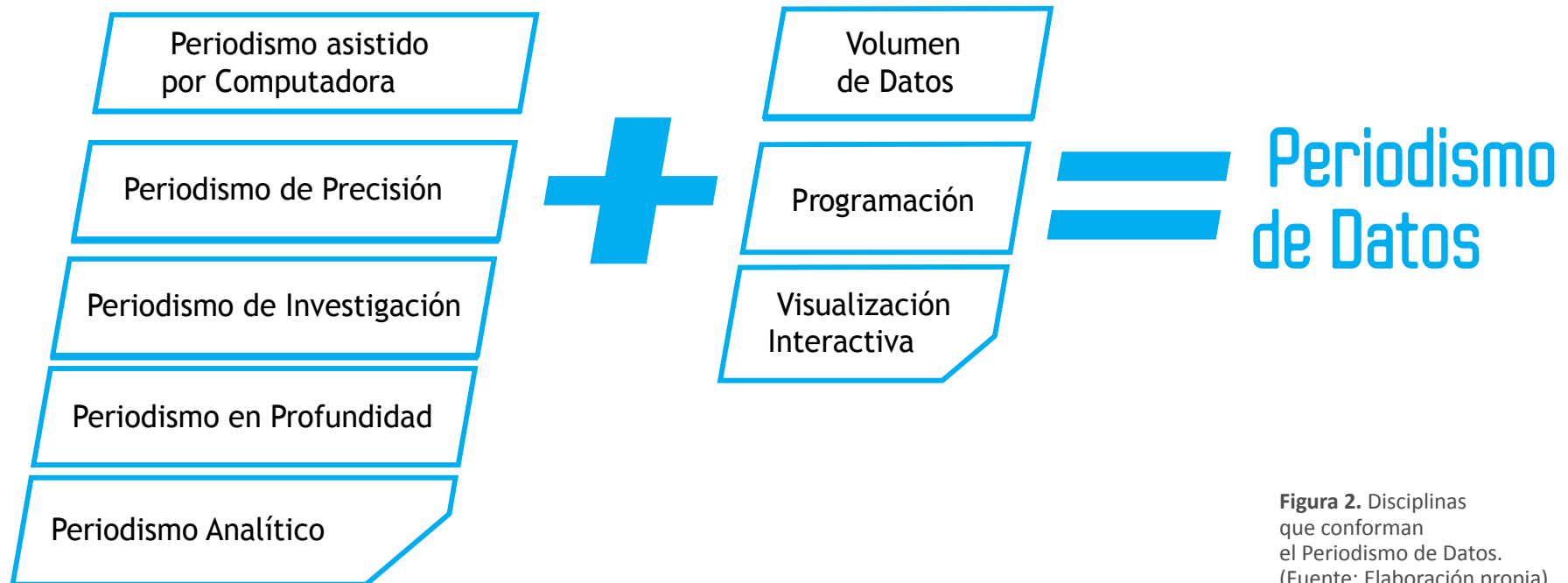


Figura 2. Disciplinas que conforman el Periodismo de Datos. (Fuente: Elaboración propia)

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. No obstante, el Periodismo de Datos, se basa en un principio fundamental del periodismo de siempre: los hechos, son sagrados” (Ramonet, 2011, pág. 18).

Sandra Crucianelli, académica de la Universidad de Texas conceptualiza al Periodismo de Datos como “una disciplina periodística que se nutre de otras muchas: de investigación, en profundidad, de precisión, asistido por computadora y analítico. En ella se trabaja con grandes volúmenes de datos, se aprovecha al máximo la visualización interactiva y se incorpora al programador al equipo periodístico” (Crucianelli, 2013, pág. 107).

Finalmente, Adolfo Antón Bravo en su artículo “El Periodismo de Datos y la web semántica” menciona que “el Periodismo de Datos se refiere al periodismo como ese lenguaje que cuenta historias de la realidad. Se utiliza mucho la expresión *los datos cuentan historias*, datos que se expresan en visualizaciones, por lo que también se habla de *narrativas visuales o visualización narrativa*. Se refieren a datos expresados de muchas formas: numéricos, alfanuméricos, contenidos textuales, bases de datos, tablas, etc. En archivos de distintos formatos” (2017, pág. 99).

Tomando como referencia las caracterizaciones mencionadas, se puede ver claramente que el P.D. no es una nueva disciplina, sino una evolución donde el periodismo tradicional se apoya de otras disciplinas y sobre todo de

herramientas tecnológicas para su ejercicio.

En síntesis, el P.D., es aquel capaz de contar historias, tomando como base grandes cantidades de datos disponibles. Para la interpretación de los datos se utilizan las clásicas habilidades de investigación de los periodistas y suma a sus capacidades, aquellos conocimientos relacionados con las herramientas informáticas, mismas que están presentes durante todo el proceso.

El P.D., es la forma en la que el periodismo se adapta a las condiciones digitales y aprovecha tanto los grandes repositorios de información digital, como la necesidad de los ciudadanos por solicitar mayor transparencia a los gobiernos.

1.3 Metodología del Periodismo de Datos

Al revisar los diferentes conceptos provenientes tanto de periodistas como académicos, salta a la vista, que el P.D. no viene a reemplazar las formas tradicionales del oficio, por el contrario, lo enriquece, aportando metodologías que son altamente cuantificables y verificables.

Giannina Segnini (2012) plantea que en el P.D. se deben contemplar 5 pasos básicos:

- **Obtención de datos.**— Es la primera fase, a simple vista pareciera ser la más sencilla del proceso. Lo cierto es que cada

vez se vuelve más compleja debido a que hay servidores públicos que no entienden por qué se solicita acceder a una base de datos - ¿para qué le puede servir a un periodista?. En este punto, la clave es no ceder e insistir con todos los protocolos que permiten el acceso a la información.

■ **Limpieza de datos.-** Estos normalmente vienen repetidos o con códigos erróneos o con información que no es de interés para el fin planteado. Entonces hay que revisarlos y estandarizarlos² de acuerdo a la necesidad. Esta es la parte más difícil para el periodista pues si no se hace con cautela, puede llegar a obstaculizar las siguientes etapas del proceso. Para que esta etapa no sea tan compleja siempre es enriquecedor tomar un curso sobre la construcción de CSV (Comma-Separated Values) o bien de manejo de hojas de cálculo.

■ **Análisis.-** Es el paso más importante en el P.D., no se trata de publicar datos, es hacer lo que los periodistas hacen de facto, sacar una conclusión. Para eso es necesario comparar la información encontrada, correlacionarla con otra, cruzar datos para descubrir tendencias, patrones y comportamientos atípicos.

■ **Verificación de la Información.-** Las bases de datos no contienen nunca la verdad

absoluta por lo que es necesario verificar cada resultado obtenido. Existen periodistas que piensan que la respuesta para toda pregunta se encuentra en Google. La mala noticia para esta clase de periodistas es que el rigor seguirá siendo la mejor arma para el periodismo de calidad.

■ **Visualización.-** Pensar como diseñador, hoy es una obligación para todo periodista. Es necesario encontrar la mejor forma de mostrar los datos a los diferentes lectores. Resulta relevante sugerir que, si se piensa la información, antes de escribirla, en términos de cómo presentarla puede ayudar a hacer informaciones más agradables y claras. La visualización entonces, se refiere a pensar y diseñar cómo presentarles a los lectores la información, que en ocasiones resulta confusa.

Adicional a los pasos mencionados, podemos agregar que, para el ejercicio del P.D. son necesarios:

- a) la investigación periodística,
- b) el conocimiento en programación y uso de software,
- c) leyes que permitan el acceso a la información.

² La estandarización de datos facilita que los datos de origen sean internamente coherentes; es decir, que cada tipo de datos tenga el mismo tipo de contenido y de formato. (Fuente: https://www.ibm.com/support/knowledgecenter/es/SSZIPZ_9.1.0/com.ibm.swg.im.iis.qs.ug.doc/topics/c_Conforming_output_data.html) (Consultado el: 9 de noviembre de 2017)

Podemos concluir que, en el desarrollo del P.D., la metodología que se aplica es iterativa, ya que en todo momento el periodista debe verificar la información, como se muestra en el siguiente diagrama:

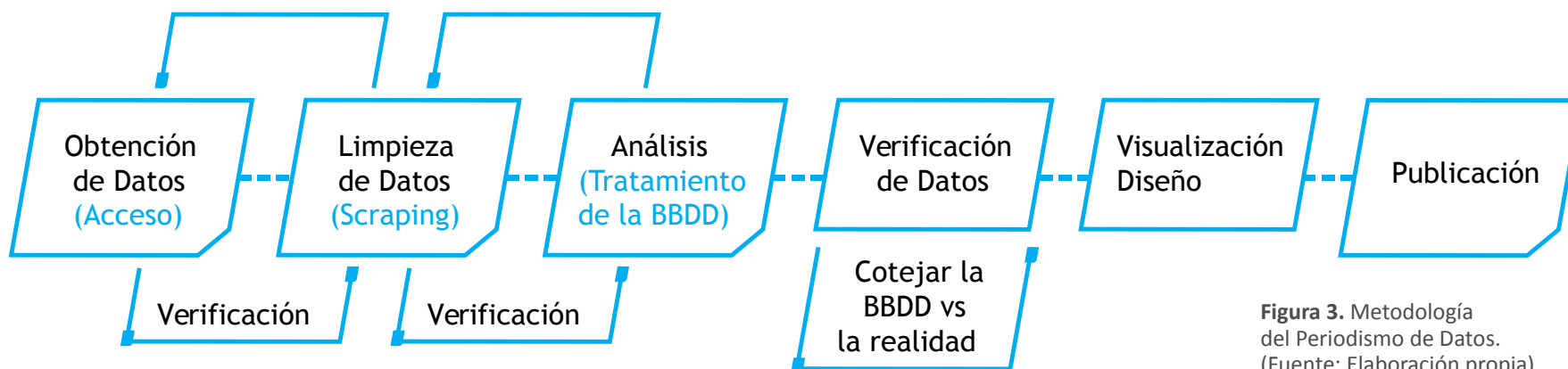


Figura 3. Metodología del Periodismo de Datos. (Fuente: Elaboración propia)

1.4 Impacto e intervención social de los nuevos medios en el periodismo

Todo sistema de información digital incide en el proceso comunicacional, afectando directa o indirectamente a los medios de comunicación y, por ende, al periodista o profesional de la noticia (Parra & Álvarez, 2004).

El periodista precisa en la actualidad de conocimientos tecnológicos de los que antes, cuando participaba exclusivamente en una parte pequeña, aunque esencial, del proceso informativo, podía prescindir. Su misión se reducía a contar historias. Es decir, a buscar noticias relacionadas

con el interés público y a presentarlas de la forma más atractiva y clara posible (Parra & Álvarez, 2004, pág. 137).

Sin embargo, hasta el momento, los periodistas se muestran reticentes a los cambios, muchas veces no se interesan ni por movilizar sus archivos al medio digital. De igual forma, hay poca predisposición a tratar dichos archivos y “explorarlos como un yacimiento minero de información”, en donde el profesional de la información se convierte en el minero que extrae, analiza y depura el contenido de dicha información (Arruti, 2001, pág. 189).

Al moverse en un mundo muy saturado de información, pensaríamos que son los periodistas, precisamente, quienes deben estar a la

vanguardia de la nueva forma de conseguir y analizar la información, es de extrañar este recelo o desidia hacia el nuevo entorno por parte de quienes, en teoría, deberían haberlo hecho suyo.

Sin tener que abordar el tema desde un punto de vista generacional, Arruti (2001), tiene una hipótesis que podría proporcionar fundamento ante tal situación:

“Desde una perspectiva periodística y desde mediados de los setenta, los medios de comunicación de los países occidentales se vieron conmovidos por una sacudida inesperada. A los periodistas les querían quitar sus entrañables máquinas de escribir y sustituírselas por unas pantallas con teclados. Evidentemente, el cambio no fue fácil. La entrada digital calentó los talleres y las redacciones a medida que se enfriaba inexorablemente el plomo en las imprentas. Hubo conflictos que marcaron la época, como los de *Fleet Street* en Londres, *The New York Times* o *The Washington Post* en Estados Unidos (eran los más reconocidos). Tras la época dorada de las grandes cabeceras, la reestructuración empresarial y la propia profesión periodística -de mano de la informática- afectó en diferentes grados casi toda la industria de los diferentes medios impresos.

Por otro lado, las áreas técnicas se han convertido en las locomotoras de la innovación tecnológica del sector, un factor clave para mantener el filo competitivo de las

empresas de comunicación. En este proceso, los periodistas han tenido casi nula participación. El departamento técnico, ha mantenido el tren de la innovación en marcha, alimentado por el carbón echado a las calderas de conocimiento tecno-informático. Los periodistas han sido meros instrumentos de sus diseños: bastaba con que aprendieran a manejar lo nuevo, no faltaba que lo entendieran. El paso de los años ha consolidado esta relación prácticamente en todos los medios de comunicación” (pág. 175).

Aunque, seguramente a los periodistas de ahora les resultaría imposible imaginar su profesión sin las computadoras, al parecer han cultivado una actitud de *tecnofobia*. La total dependencia de las máquinas, por otra parte, para cumplir con el cometido profesional ha favorecido a esta “cultura”, que se ha alzado fatalmente como una muralla a la hora de establecer el necesario puente de comunicación con el departamento técnico para transmitir las necesidades específicas de la redacción. Los técnicos, por su parte, han encontrado el terreno abonado para tratar con cierta condescendencia a quienes manejan artilugios tan complejos, versátiles y multifuncionales como si fueran meras máquinas de escribir con pantalla (Arruti, 2001).

Nunca antes se le había exigido tanto al periodista y a los profesionales de la comunicación, sobre esto Arruti (2001) enfatiza que,

“La formación de periodistas previa a su incorporación profesional y el contacto de los estudiantes con la tecnología durante sus estudios ha sido muy escasa. Los programas docentes apenas han atendido los aspectos tecnológicos. Los planes de estudio, hasta hace algunos años, no han prestado la debida atención a este problema, o lo han hecho en escasa medida. Los aspectos tecnológicos del proceso informativo, así como las herramientas de uso diario que deben utilizar los profesionales, han quedado fuera de las aulas.

En cuanto a la formación continua de los profesionales, se suele ir produciendo simultáneamente al desarrollo de su labor profesional. Más aún, la profundización de los temas a los que está ligada su labor profesional depende en muchos casos de la inquietud e iniciativa de los propios periodistas, más que de sus superiores jerárquicos profesionales” (pág. 183).

Ante tales circunstancias, ya no basta con adquirir conocimientos y habilidades tecnológicas, más o menos, actualizadas. Ahora es imprescindible que el periodista genere una notable capacidad de auto-aprendizaje. Además, el profesional de la información debe tomar una actitud que le permita afrontar los nuevos retos para lograr desarrollar un trabajo para un público, cada vez más, exigente. Si el medio se ha adaptado a los procesos de innovación tecnológica, en consecuencia, tendrían que adaptarse los periodistas también a nuevas tendencias.

Las tecnologías de los nuevos medios, tienen importantes consecuencias para los estudios de periodismo y de comunicación, que caen dentro de cuatro grandes áreas: **(1)** cómo los nuevos medios están transformando la manera de enseñar e investigar el periodismo y la comunicación, o cómo hacen los educadores su trabajo; **(2)** cómo están transformando el contenido de lo que se enseña; **(3)** cómo están transformando la estructura de las Facultades y Departamentos de periodismo y comunicación, así como las Universidades y otras instituciones de educación superior y; **(4)** cómo están transformado las relaciones entre los profesores de periodismo y sus diversos públicos, como los estudiantes, los financiadores o los competidores (Pavlik, 2005, pág. 317).

Arruti (2001) supone que,

“El conocimiento gradual de los nuevos sistemas de información, aplicado a los medios de comunicación, hará que la labor del periodista sea más eficiente y de mejor calidad, ya que pasará de ser un transmisor de hechos a convertirse en un gestor de conocimiento.

El periodista, cuyo perfil profesional se basa en el conocimiento y el uso de las aplicaciones telemáticas (como el tener correo electrónico, las notas o foros de discusión, las listas de distribución y la búsqueda puntual de información entre millones de páginas web que existen en la Red) generará incluso interactividad de información con el usuario, pues el receptor inicial del

mensaje, se convertirá en el emisor del mismo hecho al que se le han ido agregando valores y enriqueciendo el contenido durante todo el proceso de interactividad” (págs. 193-194).

Desde que el periodista investiga en la web, se está volviendo más analítico. Estará más cerca de profundizar en un hecho de modo técnico y mucho más científico, con lo que puede considerarse como material del periodismo de investigación. El periodismo y cualquier otra forma de indagación sistemática tratan de obtener una versión o interpretación de la verdad, a base de recopilar información de diversas fuentes y de reconstruir o verificar la hipótesis más probable (Pavlik, 2005, pág. 56).

Sin embargo, solo podemos esperar aproximarnos a la verdad por estimaciones a partir de los datos conocidos, para ello, el periodismo puede facilitarle al público la comprensión de un hecho o de un proceso, revelándole el máximo posible de hechos verificables, por ello, en algunos casos el *rastreo informático*, como lo denomina Arruti (2001), ha sido indispensable para la obtención de información.

“Inicialmente, el rastreo funcionó eficazmente como se demuestra en el caso del periodista George Landau, del *St. Louis Post Dispatch*, quien utilizó una computadora para comparar los registros de defunciones en Missouri, correspondiente a diez años, con las listas del censo electoral de la zona

occidental de Saint Louis. De esta forma descubrió votaciones hechas por personas fallecidas. Con una computadora pudo cruzar 500,000 registros de defunciones con más de 25,000 inscripciones del censo electoral en cuestión de horas. Sin una computadora la tarea habría sido imposible” (pág. 190).

Queda claro que la computadora se ha convertido en más que una sofisticada máquina de escribir y ahora, ante la gran cantidad de información es el software el que ahora ayuda al periodista en su proceso por gestionar y organizar. Sin embargo, solo la inteligencia humana es capaz de evaluar si algo es nuevo o importante en relación con otros datos y su contexto histórico socio-político. El periodista es quien debe ser capaz de detectar si alguna circunstancia es inconexa.

Y de aquí, podemos desprender para la reflexión, algunas consideraciones éticas. Un ejemplo de ello es la hipótesis de Pavlik (2005), quien menciona que:

“Hay una información que trata de cosas que salen de lo normal, de modo que el periodista programa un agente inteligente para hacer un recorrido por Internet en busca de anomalías en las pautas de comportamiento de los datos. Por el camino, el agente se encuentra con otro que ha sido programado para abrir ilegalmente archivos encriptados. Al observar cómo trabaja el robot, el primero de los agentes aprende a abrir archivos que

obedecen a una pauta muy poco corriente de una gran institución bancaria. Inmediatamente vuelve con la información a su amo periodista y, el periodista se abalanza tras la pista de una gran noticia sobre irregularidades financieras en una institución bancaria.

Está claro que se ha cometido un delito: un sistema informático encriptado ha sido pirateado por un agente inteligente. ¿Quién es el responsable? ¿El agente? ¿Su amo de carne y hueso? ¿Se ha violado alguna norma ética? ¿Justifica el fin (es decir, la noticia), los medios, la falta de ética? ¿Puede considerarse al periodista responsable de los actos cometidos por un agente que él no programó de manera intencionada para llevar a cabo actividades ilegales o poco éticas?” (pág. 289).

Sin duda alguna, el periodismo está sufriendo una transformación radical, lo cual está ejerciendo un efecto profundo en el papel que juega la prensa en el proceso democrático.

1.4.1 Periodismo y Hacking Cívico

En esta era tecnológica, dentro de la red, existen personajes virtuales que se autodenominan *hackers*. Contreras (2003) los define como personas apasionadas y curiosas en temas de computación, que tienen conocimientos profundos sobre

programación y desarrollan software gratuito pues, consideran que su deber es compartir proyectos y todo tipo de información con los usuarios de Internet.

Por otro lado, Raymond (2001) los describe como, personas que tienen habilidades técnicas y gran capacidad en la solución de problemas relacionados con temas de la computación.

Enfatiza que ellos construyen Internet, desde la realización de *ARPANet*, como el sistema operativo *Unix*; los hackers hacen que la World Wide Web funcione.

Los primeros en utilizar el término *hacker*, fueron un grupo de ingenieros del *Instituto de Tecnología de Massachusetts (MIT)* en el año de 1961. Año en el que se tienen los primeros indicios de la cultura *hacker* (Raymond, 2000).

Dicho grupo de *hackers* del MIT contaba con un sistema denominado “Ética del hacker”. Pau Contreras (2003) enlista los códigos que conforman esta filosofía hacker; plantea que este tema ha sido tratado por distintos autores, sin embargo, considera que Steve Levy, creador de “*Hackers: heroes of the computer revolution*”, retoma los aspectos más aplicables al movimiento de los años sesenta, por ello, enlista los siguientes puntos:

1. Toda información debe ser de libre acceso.
2. El acceso a las computadoras debe ser ilimitado y total.

3. Desconfía de la autoridad. Promueve la descentralización.
4. Un *hacker* debe ser valorado por sus *hacks*, es decir, por la calidad de sus diseños y programas, no por criterios falsos y postizos como las titulaciones académicas, la raza o la posición social.
5. Un *hacker* puede crear arte y belleza con una computadora.
6. Las computadoras pueden mejorar nuestras vidas.

Ahora bien, ¿cómo todo anterior se liga con el periodismo? Para poder entender la vinculación entre los *hackers* y el periodismo es importante retomar el caso de Wikileaks, organización que sin dudas ha venido a impulsar la generación de un nuevo tipo de periodismo (Rodríguez M. E., 2016).

Wikileaks fue fundada en 2006 por el *hacker*, *ciberactivista*, periodista y activista australiano, Julian Assange. Es gestionada por la organización *The Sunshine Press*, que publica informes y documentos en materia de interés público, teniendo como principio el anonimato de sus fuentes, exponiendo los comportamientos no éticos por parte de gobiernos, organizaciones, empresas y particulares relevantes (Moya Cantero, 2011).

Esta organización, ha desarrollado relaciones contractuales y parámetros de seguridad comunicacional no sólo con una, sino con más

de una centena de organizaciones mediáticas alrededor del planeta, así como con el medio periodístico (Rodríguez M. E., 2016). Sin lugar a dudas cambió las reglas para el periodismo. Tanto *hackers* como periodistas, así como otros involucrados, colaboran de manera conjunta en favor de que ciertos datos o información sean revelados para el beneficio de la sociedad.

Lo anterior, se debe, principalmente a que Wikileaks se rige bajo la máxima de la ética del *hacker*, y es que toda información debe ser libre y de acceso ilimitado. Sin embargo, no está claro que sea periodismo ciudadano, pese a que está elaborado por personas ajenas al mundo del periodismo tradicional, quienes ponen sus conocimientos; en este caso, informáticos, al servicio de la libre circulación de información en la red (Elías, 2015).

Gracias a esfuerzos como Wikileaks, la sociedad tiene la opción de ejercer cambios por medio de las TICs, para enaltecer sus derechos, libertades y proponer soluciones alternativas. En este contexto, las tecnologías cívicas se presentan como un conjunto de herramientas para la resolución de problemas, pero también como plataformas y dispositivos para la acción de la ciudadanía (Sánchez, Bolaños, Magallón, & Anderica, 2015).

Vivimos en una era en la que estamos rodeados de datos, cada segundo generamos datos, como en sensores aéreos, aparatos móviles, lectores de identificación de radio frecuencia, etc. (Milan & Gutiérrez, 2015).

Al ejercicio de la sociedad al trabajar con la tecnología en la manipulación de estos datos, para desarrollar propuestas en favor del bienestar de la comunidad, se le denomina *hacking cívico*, se conforma por tecnólogos y activistas en favor de la apertura de datos (Milan & Gutiérrez, 2015).

En los últimos años, las organizaciones en favor de políticas de apertura y transparencia, de los denominados Datos Abiertos han ido en crecimiento (Sánchez, Bolaños, Magallón, & Anderica, 2015). Por ello, en épocas recientes se han tenido que desarrollar nuevos retos en los que se involucra a la sociedad.

Se están generando nuevas formas de enfrentarse a los problemas de gobernanza local y global; entre ellos destaca la colaboración y participación por parte de la ciudadanía a la hora de aportar soluciones a problemas comunes. Las estructuras clásicas de gestión pública deben adaptarse a los nuevos desafíos y aprovechar las oportunidades que la nueva economía colaborativa nos está empezando a plantear (Magallón, 2014).

De esta nueva forma colaborativa, de intercambiar y gestionar conocimiento, surge el periodismo cívico, el cual, forma parte de un panorama que está creciendo en los nuevos medios. “Se trata de una forma de periodismo en la que la prensa participa de una manera activa en la vida pública de las comunidades geográficas, políticas y culturales a las que sirve” (Pavlik, 2005, pág. 12).

Pavlik (2005, pág. 338) considera que el periodismo cívico, en su forma digital no solo estimulará más la participación ciudadana en la vida pública, sino que también contribuirá en la mejora del periodismo mismo, al conectar más a los medios periodísticos locales con las instituciones y procesos que cubren.

La información y la comunicación, que se realizan a través de los medios de comunicación con el formidable apoyo de las nuevas tecnologías, tienen una importancia decisiva para el desarrollo individual y social. Son indispensables para la vida democrática, ya que para que la democracia pueda desarrollarse plenamente, la participación de los ciudadanos en los asuntos públicos, debe estar garantizada. Esta participación sería imposible si los ciudadanos no reciben la información necesaria sobre los asuntos públicos que deben procurarle los medios de comunicación (Villanueva, 2004, pág. 175).

En México, existen organizaciones que promueven este vínculo entre periodistas y *hackers*, como lo son *Codeando México*, que en el 2013 publicó en el periódico Reforma, un artículo sobre una licitación aprobada por la Cámara de Diputados de la LXII Legislatura. El artículo se centraba en una aplicación móvil con valor de 115 millones de pesos, cuyo propósito era monitorear la actividad legislativa de los diputados. En su artículo, *Codeando México*, cuestiona el costo de la aplicación, pues en la opinión de especialistas consultados, el precio de la *app* era excesivo (Tena-Espinoza-de-los-Monteros & Merlo Vega, 2017).

Este tipo de trabajos colaborativos es sin duda de vital importancia pues, para que la democracia exista, se requiere de un efectivo sistema de comunicación política, en sentido amplio, que informe y que anime a las personas a participar, haciendo que se interesen por ella. Esto se vuelve especialmente importante en la medida que las sociedades se vuelven más grandes y más complejas (McChesney p.13, 1997, citado por Pavlik, 2005).

1.5 Casos Internacionales de Periodismo de Datos

De acuerdo con el Manual de Periodismo publicado por la UNESCO, el periodismo de investigación busca 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 que dificultan su comprensión.

Es una actividad que requiere el uso de fuentes y documentos tanto públicos como privados (Kaplan, 2013).

En la actualidad cada vez son más los casos de Periodismo de Investigación con características de P.D., refiriéndonos específicamente al uso de Datos Abiertos, la aplicación de la metodología propuesta por esta nueva especialización del periodismo y la creación de historias gracias al trabajo en equipo de grupos interdisciplinarios.

A continuación, se presentan algunos casos internacionales donde se aplicó la metodología que propone el P.D. Se seleccionaron aquellos que para su desarrollo se crearon grupos de trabajo conformados por periodistas de diferentes países, quienes colaboraron de manera conjunta para la revisión de grandes cantidades de información. Asimismo, también intervinieron: programadores y expertos en *Big Data*, quienes aportaron sus conocimientos para la sistematización de la información; y de diseñadores que ayudaron a presentar la información de manera más amigable e interactiva a los lectores.



Figura 4. Portada del Reportaje Chinaleaks. (Fuente: <https://www.icij.org/blog/2014/01/highlights-chinaleaks> Consultado el: 21 de octubre de 2017)

1.5.1 Chinaleaks

El 22 de enero de 2014, numerosos medios europeos entre los que se encuentran *El País* (España), *The Guardian* (Gran Bretaña), *Le Monde* (Francia), *Süddeutsche Zeitung* (Alemania), *Asahi Shimbun* (Japón) y la Cadena BBC, publicaron una serie de informaciones denominados “Chinaleaks”³. La información era un conjunto de filtraciones que comprometían a 22,000 ciudadanos chinos, entre ellos familiares cercanos a los principales dirigentes del Partido Comunista⁴.

Estas revelaciones se realizaron por el Consorcio Internacional de Periodistas de Investigación (ICIJ) el cual destapó la trama tras las filtraciones de más de dos millones de archivos procedentes de las dos gestoras mencionadas.

La información hace referencia a las cuentas que familiares del Partido Comunista y empresarios

chinos tienen en paraísos fiscales con el fin de ocultar bienes y dinero del control oficial, pues China limita el movimiento de capital al extranjero en 50,000 dólares por habitante y año.

La iniciativa de presentar la información a través de diferentes medios alrededor del mundo el mismo día, tuvo gran impacto; a pesar de que el gobierno chino emitió la orden de bloquear a todos los medios que difundieran el caso.

Gracias al impacto que tuvo la investigación a nivel internacional, hubo varias declaraciones por parte de diferentes funcionarios del gobierno chino. Según un artículo publicado por *El País* el comisario de impuestos de China, Wang Jun, declaró que su país se comprometería a intensificar la cooperación internacional en la lucha contra la evasión fiscal (El País, 2014).

Es importante hacer énfasis en que el caso tuvo gran repercusión en China a pesar de que

³ <https://www.documentcloud.org/documents/1009191-chinaleaks-coverage.html> (Consultado el: 21 de mayo de 2017)

⁴ <https://www.icij.org/blog/2014/01/highlights-chinaleaks> (Consultado el: 21 de mayo de 2017)

fueron medios internacionales los que dieron a conocer la información. Ningún medio chino hizo eco a las filtraciones, y el gobierno bloqueó en su territorio el acceso a todas las páginas de los medios que dieron cobertura al escándalo. Fue la difusión internacional y la opinión del mundo entero la que llevó al gobierno chino a tomar acciones. Por ejemplo, el presidente Xi Jinping se comprometió a luchar contra el enriquecimiento ilícito de las elites del país.

1.5.2 Panama Papers

En 2015, una fuente anónima filtró información al periódico alemán *Süddeutsche Zeitung*, la filtración contenía los datos de clientes que evadían impuestos en sus países a través de paraísos fiscales con la ayuda del consorcio de abogados panameño *Mossack Fonseca*. En la lista se encontraban jefes de estado, políticos, personalidades del deporte, arte, finanzas, entre otros.

Süddeutsche Zeitung compartió con el Consorcio Internacional de Periodistas de Investigación (ICIJ) 2.6 terabytes de información los cuales contenían 11.5 millones de documentos, la organización planteó una nueva forma de realizar la investigación de manera global con un equipo de 370 periodistas repartidos en 80 países pertenecientes a más de 100 medios de comunicación (Obermaier & Obermayer, 2016).

Mediante el uso de herramientas que ayudaban



Figura 5. Periodistas entrevistando a Carlos Sousa, Dir. de RRPP de Mossack Fonseca & Co. (Fuente: panamapapers.icij.org. Consultado el: 7 de noviembre de 2017)

a la sistematización de los datos, como el buscador *Blacklight* (permite identificar palabras clave y ubicarlas en archivos). Así como el de una red social encriptada, (a la cual obtenían acceso únicamente los investigadores) donde compartían hallazgos y realizaban videoconferencias para plantear estrategias y definir detalles de la publicación.

Los miembros del equipo eran conscientes de que para que la investigación fuera trascendente, era necesario lograr impactar a los lectores; por ello decidieron generar alianzas con diferentes medios internacionales para publicar el artículo el mismo día a la misma hora. Tal hecho, provocó que la noticia fuese tema central de discusiones durante varios días, marcando así una nueva era para el periodismo, un nuevo tipo de periodismo basado en filtraciones de gran cuantía que ha traído a la luz la necesidad de compartir esfuerzos y nuevas formas de colaboración periodística en redes globales (Rodríguez E. M., 2016).

1. 6 Resumen del capítulo 1

En este capítulo se presentó la importancia del periodismo como una actividad que se encarga de recopilar, analizar, interpretar y comunicar información relacionada con acontecimientos de interés social. De igual forma, se reconoció la relevancia de la tecnología, que actualmente se encuentra vinculada a prácticamente todos los ámbitos de la vida cotidiana.

De igual manera, se ha explorado con base en investigaciones previas, lo que sucede cuando el periodista utiliza esta tecnología al posibilitar el procesamiento de información a través de medios artificiales como son las computadoras no solo para comunicarse, buscar y almacenar sus datos, sino para contar historias a partir de grandes cantidades de datos.

Esto deriva en el fenómeno llamado P.D. como una evolución donde el periodismo tradicional se apoya sobre todo en herramientas tecnológicas para su ejercicio pero además en otras disciplinas. Si bien es cierto que para la interpretación de los datos se continúan utilizando las clásicas habilidades de investigación de los periodistas y son ellos quienes deben ser capaces de detectar si alguna circunstancia es inconexa, las herramientas informáticas están presentes durante todo el proceso.

Para ello, Giannina Segnini plantea la contemplación de cinco pasos básicos:

- Obtención de datos
- Limpieza de datos
- Análisis
- Verificación de la información
- Visualización

Las herramientas utilizadas en cada uno de estos pasos implican la suma a las capacidades del periodista con aquellos conocimientos relacionados con ellas o bien, de la incorporación de programadores y diseñadores al equipo de trabajo, ahora interdisciplinario.

A su vez, el P.D. genera impacto en otras áreas, pues además, de dar un uso a los grandes repositorios de información digital disponibles, el P.D. aprovecha también la necesidad de las sociedades por solicitar mayor transparencia a sus gobiernos, favoreciendo también el desarrollo de sociedades más democráticas.

La participación de los ciudadanos en los asuntos públicos, debe estar garantizada pero sería imposible si los ciudadanos no reciben la información necesaria sobre los asuntos públicos que deben procurarles los medios de comunicación, por lo que el apoyo de las nuevas tecnologías es indispensable para la vida democrática.

Hasta aquí, hemos revisado temas que sin duda, nos hacen destacar la importancia del trabajo colaborativo. De esta nueva forma colaborativa

de compartir y gestionar conocimiento es que encontramos al periodismo cívico, que con la participación de los ciudadanos busca acortar la distancia entre estos y sus gobiernos. También, encontramos el hacking cívico, con el que la sociedad tiene la opción de ejercer cambios por medio de las TICs, y que también nos hace plantearnos ciertas consideraciones éticas. Se presentaron casos internacionales en donde el desarrollo de la metodología propuesta por el P.D. implicó la creación de grupos de trabajo conformados por periodistas de diferentes países, con el fin de colaborar en la revisión de grandes cantidades de información. Tras estos esfuerzos, por supuesto, están también incluidos programadores, expertos en Big Data y diseñadores.

Tomando como referencia los estudios encontrados realizados a la fecha, los medios observados y los casos de éxito internacionales explorados: WikiLeaks, Chinaleaks y Panama papers, en general, se puede concluir que:

- El periodista precisa, en la actualidad, de conocimientos tecnológicos de los que antes, podía prescindir. Sin embargo, tales conocimientos que incluyen herramientas computacionales o conocimientos muy técnicos y que son utilizados por los profesionales del P.D. aún no están contemplados en los planes de estudio para la formación de recursos humanos. Por lo tanto, ahora es imprescindible que el periodista genere una notable capacidad de autoaprendizaje e incluso de aprendizaje colaborativo.

- El P.D. no viene a reemplazar las formas tradicionales del oficio, sino que lo enriquece aportando metodologías que son altamente cuantificables y verificables, lo que además aportará una mayor credibilidad a la investigación.

- En el desarrollo del P.D., la metodología que se aplica es iterativa, ya que en todo momento el periodista debe verificar la información.

- Adicional a los pasos que contempla su metodología, antes mencionados, para el ejercicio del P.D. son necesarios: **a)** la investigación periodística, **b)** el conocimiento en programación y uso de software y **c)** leyes que permitan el acceso a la información.

- Son cada vez más los casos de Periodismo de Investigación con características de P.D., refiriéndonos específicamente al uso de datos abiertos y “en los últimos años, las organizaciones en favor de políticas de apertura y transparencia, de los denominados datos abiertos han ido en crecimiento” (Sánchez, Bolaños, Magallón, & Anderica, 2015).

En el siguiente capítulo se realiza un análisis de la situación actual del Periodismo en México, además se presentan algunos ejemplos de investigaciones en donde grupos interdisciplinarios de periodistas, programadores y diseñadores mexicanos han aplicado la metodología propuesta por el P.D.

Periodismo de Datos en México

Capítulo **2**

2.1 Periodismo en México

Se podría creer que la tarea primordial del periodismo moderno es la de informar, interpretar y guiar a la sociedad de la que es parte, lamentablemente en nuestro país eso no se cumple. En México, la realidad es un poco más compleja; para algunos la principal función del periodismo es la política, no existen la objetividad ni mucho menos la imparcialidad periodística, en México el periodismo es por sobre todas las cosas, una forma de hacer política (Lizana, 2006).

Al final del siglo XX, los medios de comunicación se han convertido en el centro de poder contemporáneo de nuestra nación. El periodismo mexicano ha pagado muy caro los avances de la libertad de expresión, en la última década ha sido considerada una profesión de alta peligrosidad y la lista de periodistas encarcelados, perseguidos y presionados no ha cesado de incrementarse (Fabila, 2007).

De 2000 a la fecha, *Artículo 19* ha documentado el asesinato de 111 periodistas en México, en posible relación con su labor periodística. Del total, 103 son hombres y 8 son mujeres (Artículo 19, 2017).

Según el reporte de 2016 realizado por Freedom House, organización no gubernamental dedicada a la expansión de la libertad y la democracia alrededor del mundo; menciona que México es uno de los países más peligrosos para ejercer como periodismo, pues la libertad de prensa se encuentra siempre en constante amenaza.

Los periodistas se enfrentan constantemente al acoso, la intimidación, los ataques físicos y la autocensura, principalmente en aquellas zonas donde el narcotráfico posee mayor poderío (Freedom House, s.f.).

En el reporte de 2016 sobre México, Freedom House, resume aspectos tanto legales, políticos y económicos que impiden que en nuestro país se ejerza un periodismo libre, crítico y sin censura. Desafortunadamente, las leyes actuales están diseñadas para darle mayor poder a los grandes grupos de medios favoreciendo actividades monopólicas. No existen leyes que protejan a los periodistas que arriesgan su vida realizando investigaciones de interés nacional, por el contrario, se busca a toda costa censurar a los que se atreven a cuestionar a aquellos que se encuentran en el poder.

Muchos de los periodistas no escriben para consumo de los lectores, para orientar a la sociedad, sino para que puedan ser leídos por los sujetos del poder y sus colegas de mayor rango en el proceso informativo. Por ello, el mexicano promedio adquiere información precaria a través de medios como la televisión o la radio.

Actualmente, México se posiciona como un país peligroso, pues los crímenes contra la libertad de expresión y contra el derecho que tenemos como ciudadanos a estar informados se han venido incrementando, principalmente hacia los que ejercen la labor social de informar a través del periodismo (Díaz, 2017).

“El periodista es un factor importante en el

proceso informativo y su ética profesional está orientada al desempeño correcto de su labor, así como contribuir a la reducción o eliminación de deformaciones” (Gómez, 2008, pág. 54).

Como miembros de la sociedad, al realizar una tarea de tal seriedad para la misma, se espera que los integrantes de los medios informativos y en concreto, de los periodistas, tengan los niveles de ética que les permitan el leal desempeño de su labor; pero la realidad a la que se enfrentan es muy diferente a este ideal. Y es que, desde hace tiempo, los agresores o aparentes responsables de vulnerar las libertades de información y expresión más recurrentes han sido personas del poder público en todas sus variantes. Más ahora, se suman otros actores como grupos sociales o incluso el crimen organizado (Félix Anduaga, Alvarado Noriega, & Mendivil Alvarado, 2015).

Para Gómez (2008, pág. 71), “las amenazas, agresiones físicas, el hostigamiento y los atentados a los periodistas y medios, sigue siendo una de las prácticas más comunes para amedrentar la libertad de expresión y el derecho a la información de periodistas y medios de comunicación. La autocensura se ha convertido en una consecuencia del temor y una forma para la sobrevivencia”.

Así mismo, Gómez (2008) manifiesta también el *bloqueo informativo* u *obstaculización* como una agresión indirecta y más discreta en el contexto de los periodistas, pues dificulta el proceso de su labor.

Fernández (2004) menciona sobre el periodismo que,

El “deber ser” de este oficio necesita ser repensado al tiempo que se construye el pacto legal del acceso a la información pública. Hay una relación directa entre ambos momentos. Las asociaciones de periodistas o la actitud individual de quienes ejercen este oficio deben a la sociedad códigos de ética que expresen su compromiso en valores, al mismo momento que los otros actores sociales levantan la estructura legal cuyo pacto pueda ser cumplido por todos (p. 13).

Para Ruiz (2006)

“La ley Federal de Transparencia y Acceso a la Información Pública Gubernamental (LFTAIPEG) ha modificado profundamente la vida pública de México, y quizás apenas podemos comenzar a dimensionar adecuadamente el impacto de su promulgación. Acceder a la información del gobierno federal ha sido una herramienta eficaz para entender la gestión del gobierno a través del escrutinio directo de archivos e información diversa; para activar mecanismos anticorrupción; para empoderar a la ciudadanía como controlador de las actividades del gobierno y para disminuir la discrecionalidad de las decisiones del gobierno al ponerlas al escrutinio público” (p.15).

Sobre el derecho a la información, Carpizo y Villanueva (2004) sostienen que en un sentido amplio y de acuerdo con el artículo 19 de la Declaración Universal de los Derechos Humanos,

■ Es la garantía que toda persona posee a atraerse información, a informar y a ser informada.

■ De la definición apuntada se desprenden los tres aspectos más importantes que comprende dicha garantía fundamental:

- a) El derecho a atraerse información;
- b) El derecho a informar, y
- c) El derecho a ser informado.
- d) El derecho a atraerse información incluye las facultades de:
 - i) acceso a los archivos, registros y documentos públicos; y,
 - ii) la decisión de qué medio se lee, se escucha o se contempla.
- e) El derecho a informar incluye las
 - i) libertades de expresión y de imprenta; y
 - ii) el de constitución de sociedades y empresas informativas.
- f) El derecho a ser informado incluye las facultades de
 - i) recibir información objetiva y oportuna;
 - ii) la cual debe ser completa, es decir, el derecho a enterarse de todas las noticias; y
 - iii) con carácter universal, o sea, que la

información es para todas las personas sin exclusión alguna (Escobar de la Serna, p. 54-96 1984, citado por Villanueva 2004).

“La información debe entenderse en un sentido amplio que comprende los procedimientos – acopiar, almacenar, tratar, difundir, recibir – hechos, noticias, datos, opiniones, ideas –; y sus diversas funciones” (López, p. 176, citado por Villanueva 2004, pág. 20).

Por otro lado, Guerrero (2008, pág. 325) comenta,

“En cualquier sociedad, el derecho a la información y el derecho a la privacidad tienen alcances y límites que se confunden. Respetar el derecho a la privacidad implica no comunicar a los demás lo que compete únicamente a los individuos. Al mismo tiempo, el derecho a la información es el derecho a saber de los demás: autoridades, gobernantes, beneficiados, regulados, reguladores, servidores públicos, ciudadanos, siempre se trata de la información sobre la condición y el comportamiento de personas. La discusión ha sido inevitable entre otras cosas porque se trata de valores en tensión. Las nociones de intimidad y privacidad son subjetivas: dependen de consideraciones culturales y sociales, derivan de la perspectiva particular del individuo y su relación con sus semejantes y entorno.

Los límites de la intimidad y la privacidad se plantean siempre con relación a los demás y en un momento dado. De ahí la dificultad de plantear propuestas generales. De ahí también el gran reto actual frente al desarrollo de las nuevas tecnologías de comunicación, que facilitan enormemente los intercambios personales y dejan registro de ello. Esos avances tecnológicos, con sus indiscutibles ventajas, han venido a modificar significativamente el contexto de la comunicación y las relaciones entre las personas” (pág. 325).

Gómez (2008, pág. 9) menciona como una apuesta por los estudiosos del Derecho de la Información en su doble contenido: libertad de expresión y transparencia y acceso a la información. Además de ser congruentes en la presentación de resultados transparentando la metodología, lo que permite finalmente seguir la ruta crítica a otros estudiosos a efecto de que puedan constatar los datos y enriquecer el análisis que del mismo se puede llevar a cabo.

El derecho de acceso a la información pública no es solo combatir la corrupción, lograr la rendición de cuentas y transparencia, es una herramienta para mejorar la calidad de vida de las personas. La toma de decisiones, ya sea individual o colectiva, parte de la información con que se cuenta. Si la información obtenida es reducida o equivocada, las decisiones serán similares. Por el contrario, un ciudadano mejor informado, será capaz de tomar mejores decisiones.

La educación nuevamente tiene un papel importante en este proceso, pues las personas deben conocer la información a la que tienen derecho, así como lo que pueden hacer con tal información una vez en sus manos.

El México contemporáneo requiere de una prensa que esté dirigida a la sociedad civil, a los lectores y no al poder. De una prensa con un código de ética y una moral sólida. La combinación de estas permitirá al ejercicio periodístico imponerse a la censura ejercida tanto por los dueños de los medios como por el gobierno (Cazaux, 2016).

2.2 Periodismo de Datos en México

Como se mencionó anteriormente el P.D. llegó a España y Latinoamérica hasta los años 90s, gracias a la traducción del libro de Meyer (Periodismo de Precisión, nuevas fronteras de la investigación periodística) al castellano. Desgraciadamente nunca vino acompañada del apoyo institucional de los medios para operar de manera articulada investigaciones con bases de datos. Mientras que en Estados Unidos importantes redacciones fueron capaces de incorporar equipos de programadores, especialistas en estadística y visualización de datos, en el mundo hispanoparlante los esfuerzos fueron aislados (Ávila, 2013). En nuestro país no existen muchos medios

que hayan incorporado a su redacción un equipo para hacer P.D. Se han dado casos aislados como el equipo que conforma “El Universal Data”, el equipo que colaboró en el proyecto “Narco Data”⁵ de Animal Político o bien el extinto equipo que integraba “Milenio Data Lab”⁶. Este último fue desarmado cuando se dio una incompatibilidad editorial entre el equipo de colaboradores y los dueños del periódico. También existen esfuerzos por la vía independiente, trabajos como el del sitio “¿Quién compró?”⁷, lugar donde se publican artículos generados con bases de datos alimentadas por facturas del Senado. Estos ejemplos son muestra de que en México se están dando pequeños pasos hacia un ejercicio periodístico que trabaje con bases de Datos Abiertos y en pro del P.D.

El caso de México, es particularmente curioso, existe la Ley de transparencia que entró en vigor en 2003⁸, sin embargo, parece estar escrita de tal

manera en que sea difícil obtener información. Normalmente se requieren una serie de aprendizajes para poder solicitar información vía transparencia con resultados exitosos.

Además de los pocos o nulos esfuerzos que están llevando a cabo los medios tradicionales para generar P.D., existen organizaciones como “SocialTIC” y “Escuela de Datos”, que enfocan sus esfuerzos a impartir capacitaciones a periodistas para el manejo de bases de datos, así como generar proyectos colaborativos.

Analizando el contexto de los periodistas mexicanos, es evidente que no la tienen fácil. Nuestro país es considerado el más peligroso de América Latina para los periodistas (Forbes, 2015). Los periodistas mexicanos han tenido que enfrentar los ataques del crimen organizado y de las propias estructuras del poder, por lo que en muchas ocasiones sortear la violencia ha sido prioritario.



Figura 6. Twitter del medio ¿Quién Compró?. (Fuente: <https://twitter.com/QuienCompró> Consultado el: 2 de noviembre de 2017)

⁵ <http://narcodata.animalpolitico.com/> (Consultado el: 12 de octubre de 2017)

⁶ <http://www.milenio.com/datalab/> (Consultado el: 5 de octubre de 2017)

⁷ <https://twitter.com/quien-compro?lang=es> (Consultado el: 1 de octubre de 2017)

⁸ <https://www.gob.mx/conapesca/documentos/ley-de-transparencia> (Consultado el: 21 de octubre de 2017)

2.3 Ejemplos de Periodismo de Datos en México

A continuación se presentan los casos más relevantes de P.D. en México; para definir los casos que se presentan en esta sección del capítulo, se identificaron aquellas investigaciones realizadas por equipos interdisciplinarios, mencionados anteriormente, y que además utilizaron datos abiertos.

2.3.1 La Estafa Maestra

La primera semana de septiembre de 2017, Mexicanos contra la Corrupción y la Impunidad (MCCI)⁹, asociación civil dedicada a combatir la corrupción y Animal Político¹⁰, portal digital centrado en la información política, presentaron una investigación dando seguimiento a un reporte realizado por la Auditoría Superior de la Federación (ASF), donde se destacaba el posible desvío de recursos por parte de dependencias gubernamentales hacia empresas irregulares. En el análisis de la Cuenta Pública 2013¹¹ se destacaba la participación de universidades públicas en desvíos de recursos federales.

Tomando como referencia los análisis de la ASF, realizar más de 500 solicitudes de información, revisar el estatus legal de las empresas que recibieron recursos públicos, visitar oficinas,

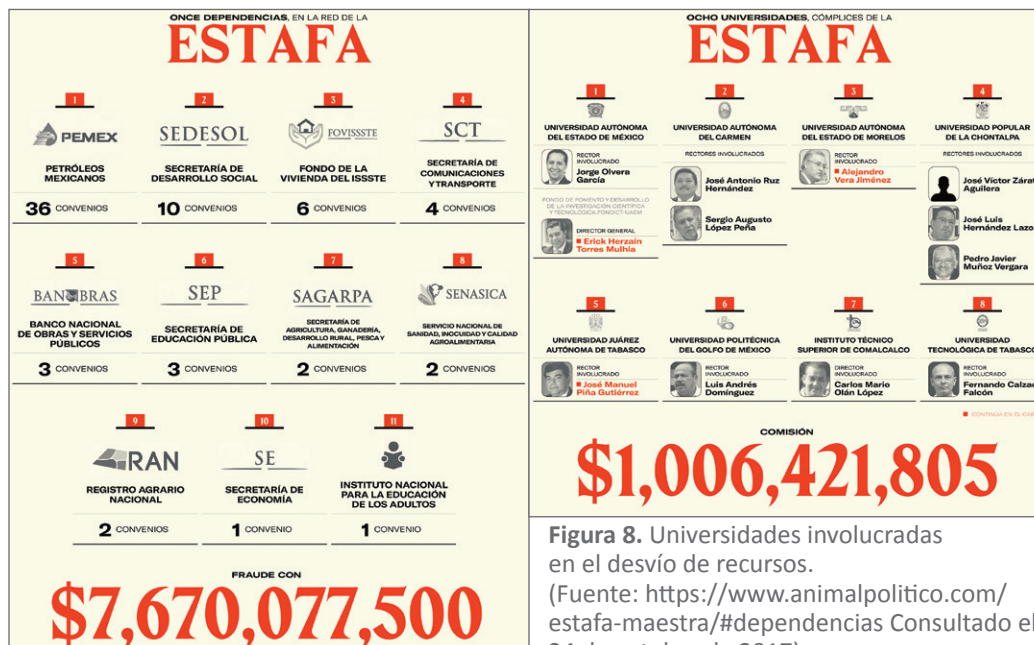


Figura 7. Dependencias gubernamentales involucradas. (Fuente: <http://www.animalpolitico.com/estafa-maestra/#dependencias> Consultado el: 22 de octubre de 2017)

Figura 8. Universidades involucradas en el desvío de recursos. (Fuente: <https://www.animalpolitico.com/estafa-maestra/#dependencias> Consultado el: 24 de octubre de 2017)

cotejar información legal y fiscal, así como entrevistar a los “propietarios”; los periodistas crearon una base de datos con más de 13 mil registros que dan cuenta sobre las características de cada una de las empresas subcontratadas. La base de datos derivada de la recopilación de la información resume 73 convenios revisados, 187 empresas investigadas y 7 mil 670 millones de pesos entregados en contratos ilegales.

“La Estafa Maestra”¹², fue el título que recibió la investigación, aludiendo al exgobernador del estado de Veracruz, Javier Duarte, quien durante

⁹ <http://contralacorrupcion.mx> (Consultado el: 24 de octubre de 2017)

¹⁰ <https://www.animalpolitico.com/> (Consultado el: 24 de octubre de 2017)

¹¹ <http://www.asf.gob.mx/Trans/Informes/IR2013i/Paginas/Master.htm> (Consultado el: 24 de octubre de 2017)

¹² <https://www.animalpolitico.com/estafa-maestra/index.html> (Consultado el: 24 de octubre de 2017)

su mandato entregó contratos a empresas inexistentes para justificar el gasto del presupuesto del estado, dinero que en realidad terminaba en manos de los funcionarios de su gobierno. Duarte, en comparación con el gobierno federal quedaba como un principiante en el desvío de recursos públicos, pues para hacer más difícil el rastreo de los desvíos no asignaba los contratos de manera directa a las empresas; su estrategia –un poco más compleja– consistía en entregar los recursos directamente a las universidades, para que fuesen ellas quienes se encargaran de la contratación de las empresas irregulares, no sin antes cobrar una millonaria comisión –aún sin prestar servicios–.

Según la investigación de los 7 mil 670 millones de pesos asignados a 8 universidades para la supuesta prestación de servicios, “al menos 3 mil 433 millones de pesos simplemente desaparecieron en empresas ilegales; otros mil millones, se convirtieron en comisiones para las universidades y el resto se convirtió en servicios contratados a precios inflados” (Roldán, Castillo, & Ureste, 2017).

Publicaciones en diversos medios de comunicación dieron difusión y seguimiento a las irregularidades, sin embargo, no ha sido suficiente para que la Procuraduría General de la República (PGR) informe sobre el resultado de las investigaciones, aun cuando dependencias como la Secretaría de Educación Pública se han sumado a estas denuncias.

“Después de la publicación, funcionarios públicos de las instancias involucradas aseguraron en diferentes entrevistas que habían

resuelto las observaciones hechas por la ASF, por las irregularidades detectadas en los convenios y contratos.

Una revisión al sistema público de la ASF arrojó que estas dependencias y universidades, involucradas en La Estafa Maestra, tienen pendientes de resolver 112 observaciones, que suman 3 mil 947 millones de pesos” (Roldán, Castillo, & Ureste, 2017).

No es la primera vez que la ASF detecta el desvío de recursos durante su proceso de realización de informes, como tampoco lo es el incurrimento de algunas universidades en tales faltas. Algunos ejemplos que dan cuenta de ello, son los reportajes de la revista Proceso: “Desnuda ASF fraude masivo en Cruzada contra el Hambre¹³” y “Reincide Universidad de Morelos en desvío millonario en fondos de la Secretaría de Desarrollo Social (Sedesol)¹⁴”. En ambos reportajes se afirma que existieron convenios de la Sedesol y las Universidades del Estado de México y Morelos con el objetivo de encubrir un fraude masivo detrás de la “*Cruzada Nacional contra el Hambre*”.

2.3.2 El (falso) éxito de la Cruzada contra el Hambre

Esta investigación sin lugar a dudas resulta significativa no solo porque aporta una cantidad importante de información sobre el tema que se aborda, además involucra al “*Grupo Editorial*

¹³ Disponible en: <http://www.proceso.com.mx/396327/la-cruzada-nacional-contra-el-hambre-un-fraude-masivo> (Consultado el: 24 de octubre de 2017)

¹⁴ Disponible en: <http://www.proceso.com.mx/430529/reincide-universidad-de-morelos-en-desvio-millonario-de-fondos-de-la-sedesol> (Consultado el: 24 de octubre de 2017)

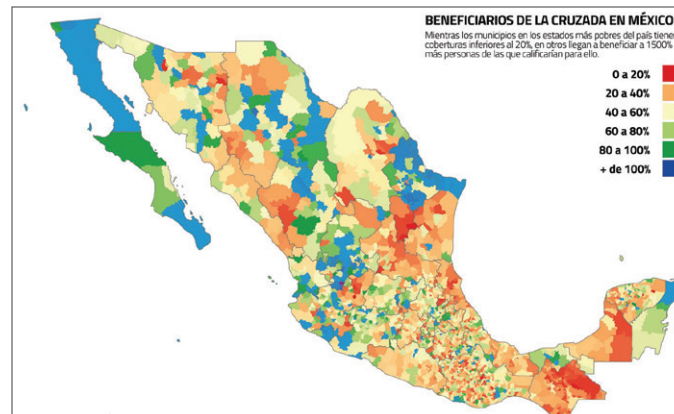


Figura 9. Visualización de “El (falso) éxito de la Cruzada contra el Hambre”. (Fuente: http://www.milenio.com/datalab/falso-exito-Cruzada-Nacional-Hambre_0_696530646.html Consultado el: 24 de octubre de 2017)

Milenio”. Uno de los primeros medios de comunicación en el país en lanzar una iniciativa en pro del ejercicio del P.D.

El ahora extinto colectivo “*Milenio Data Lab*”, fue un grupo interdisciplinario dedicado en su totalidad al desarrollo de investigaciones de temas diversos, basadas en datos abiertos. Su investigación más controversial es la que se resume a continuación. La “*Cruzada Nacional contra el Hambre*” (CNCh), el programa social más importante del sexenio fue presentado el 21 de enero de 2013. La CNCh prometía ser una estrategia de política social con el objetivo de solucionar el grave problema que representa el hambre en México. El lunes 7 de marzo de 2016, se publica esta investigación, misma que había llevado alrededor de cinco meses solo en procesar “300 mil bases independientes oficiales y públicas de beneficiarios de la estrategia social a nivel localidad y que fueron sistematizados en una base de datos que

permitiera entender lo que sucede en las entrañas estadísticas de la *Cruzada*” (Milenio Data Lab, 2016).

Después del análisis de la información recopilada, los datos dejaron al descubierto que la estrategia contra el hambre con mayor cobertura en todo el país, había distribuido de manera desigual los recursos en las diferentes entidades federativas y no atendía a aquellos estados donde la población vulnerable es mayor.

En el portal, podía visualizarse un mapa en donde por un lado, en lugares donde la pobreza es extrema, la cobertura era nula o menor de 40% y por otro lado, en algunas entidades se apreciaba un exceso de cobertura, como se muestra en la Figura 9. Otro hallazgo interesante en los datos, fue que muchos de los contabilizados como “*beneficiarios*”, recibieron apoyo de otro tipo de programas que, si bien mejoraban de alguna forma su condición, no se combatía realmente el problema del hambre (Milenio Data Lab, 2016).

Pese a las instancias e instrumentos creados para instaurar la CNCh, el análisis de los datos llevó a identificar la ausencia de una metodología establecida previa a la distribución de recursos; es decir, no se definió de manera puntual las características de los “beneficiarios”, quienes se suponía eran personas en situación real de carencia alimentaria. Tampoco se establecieron los mecanismos de evaluación del impacto de cada uno de los programas derivados de la *Cruzada*, además, la publicación sostenía que en el portal oficial de la CNCh, se publicaban datos falsos donde se reportaba una cantidad mayor de personas beneficiadas a las comunicadas en sus mismos datos oficiales (Milenio Data Lab, 2016).

El día siguiente a la publicación de la investigación, el director de *Milenio*, Carlos Marín recibió la visita de la ahora extitular de la Sedesol, Rosario Robles, así como la de otros funcionarios. La visita, tuvo como repercusión la censura de la investigación, la cual fue desmentida con el texto “Cruzada antihambre sí redujo carencias: INEGI¹⁵”.

A pesar de ser desmentida, la investigación continuó disponible en línea, no sin antes sufrir algunas modificaciones en su título pasando inicialmente por “El éxito de la Cruzada contra el Hambre” al más reciente “Los números de la Cruzada contra el Hambre”. Además de los cambios de nombre, hubo cambios significativos, como la renuncia de la periodista *Karen Cota Soto*, responsable de la investigación y gradualmente la del resto del equipo del colectivo que conformaba en sus inicios *Milenio Data Lab*.

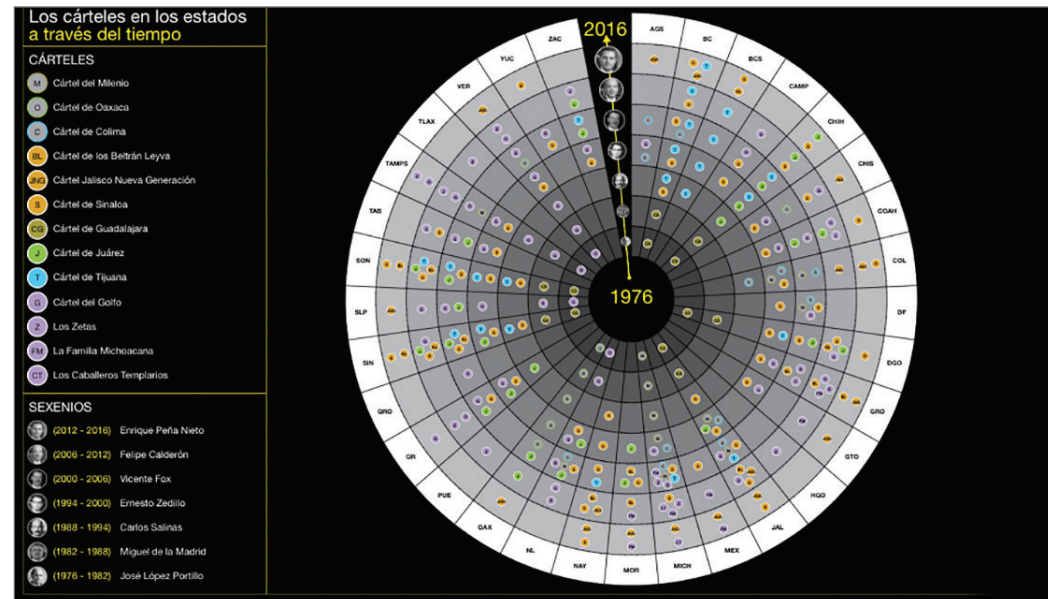


Figura 10. Visualización de los cárteles en los estados a través del tiempo. (Fuente: <http://narcodata.animalpolitico.com/> Consultado el: 25 de octubre de 2017)

2.3.3 Narcodata

En octubre de 2015, *Animal Político* en conjunto con *Poderopedia* lanzaron una plataforma digital que integra una radiografía interactiva y reportajes especiales sobre las transformaciones que han tenido las diversas organizaciones delictivas (dedicadas principalmente al narcotráfico) en los últimos cuarenta años en México.

La investigación se sustenta en la recopilación de información obtenida de la Procuraduría General de la República, a través de la Ley de Transparencia y Acceso a la Información; así como revisiones bibliográficas, hemerográficas, entrevistas y la asesoría de algunos expertos en temas de seguridad.

¹⁵ Disponible en: http://www.milenio.com/politica/Cruzada_antihambre_redujo_carencias-Cruzada_Nacional_contra_el_Hambre_0_697730249.html (Consultado el: 24 de octubre de 2017)

Narcodata, pretende explicar mediante un mapa interactivo del narcotráfico en México, el surgimiento de los cárteles de la droga, sus líderes, los conflictos entre ellos, su expansión geográfica, sus principales actividades delictivas y las conexiones entre dichas organizaciones y el poder (Animal Político, 2015).

NarcoData un proyecto en curso, ya que pretende continuar actualizando la información sobre el tema. Un punto a destacar es que en la plataforma, se encuentran disponibles las bases de datos que alimentan sus visualizaciones, esto permite a los lectores explorar y someter la información a un análisis.

2.3.4 Desaparecidos

Las cifras oficiales no revelan la realidad de los integrantes que han perdido familias tanto mexicanas como colombianas, víctimas del crimen organizado y la trata de personas, un contexto cotidiano que México y Colombia comparten (Figura 11). De esta situación es que surge la investigación “Desaparecidos”, misma que se propuso “contar el número de personas desaparecidas en cada uno de los países a lo largo de la última década, mostrar el incremento estadístico en los últimos años y principalmente, narrar las historias de cómo desaparecieron” (El Universal Data, 2015).

En México, el incremento en la cantidad de personas desaparecidas desde 2016, es atribuido

principalmente a la situación de violencia desmedida que trajo consigo la estrategia contra el crimen organizado que decidió emprender el gobierno del ahora expresidente Felipe Calderón Hinojosa (2006-2012). La información recopilada para generar la base de datos sobre México se obtuvo mediante el Registro Nacional de Personas Extraviadas o Desaparecidas (RNEPD) del gobierno federal, los registros más antiguos que se tienen en su base es de 1968; sin embargo, la base de datos recopila únicamente aquellas denuncias previas existentes en las procuradurías locales, es decir, la información con la que se cuenta es incompleta, pues existen casos que no están documentados.

En el caso de Colombia, se obtuvieron los datos del Registro Nacional, que es una base de datos del Instituto de Medicina Legal. Su base de datos cuenta con registro de las denuncias realizadas a

Figura 11. Visualización de “Desaparecidos”. (Fuente: <http://interactivo.eluniversal.com.mx/desaparecidos/mexico.html> Consultado el: 24 de octubre de 2017)



partir de 1938. Un hallazgo interesante dentro de los datos es que, las desapariciones llevadas a cabo en su territorio entre 2007-2008, se dieron lugar en el municipio de Soacha donde impera la pobreza, inseguridad y guerra (El Universal Data, 2015).

Las bases de datos utilizadas en el proyecto, están disponibles para su descarga y a su vez, es un proyecto abierto a la información que pueda mejorar sus datos.

Fue considerado como el mejor trabajo publicado en español en 2015 y recibió el Premio Ortega y Gasset concedido por el diario El País.

El proyecto promete continuar actualizando sus cifras, pero a la fecha no se conocen publicaciones posteriores a la entrega de 2015.

Los casos presentados son un claro ejemplo de la relevancia que tiene el ejercicio del P.D. en un país como el nuestro, pues son una abstracción representativa de los temas sociales que preocupan a una sociedad como la nuestra y que, al mismo tiempo se desean cambiar.

Resulta importante destacar que, sin bien los casos presentados no han dado señales de que se haya dado seguimiento por parte de las autoridades correspondientes, son una muestra de cómo los datos que está liberando el gobierno (D.A.) pueden ser el sustento perfecto para realizar este tipo de investigaciones. La información está disponible, sin embargo, es necesario que existan más profesionales del periodismo que tengan los conocimientos para aprovecharla y darle seguimiento.

Además de los escasos esfuerzos llevados a cabo por los medios tradicionales para generar P.D., existen organizaciones como Escuela de Datos, Chicas Poderosas, SocialTIC y el Laboratorio para la Ciudad, que han desarrollado eventos como *hackatones* o retos tecnológicos para realizar propuestas que hagan uso de datos abiertos en México y denuncien temas, como el presupuesto gubernamental o bien, generen propuestas de carácter ciudadano.

“SocialTIC” y “Escuela de Datos”, enfocan también sus esfuerzos a impartir capacitaciones a periodistas para el manejo de bases de datos, así como generar proyectos colaborativos.

Este tipo de trabajos colaborativos es, sin dudas, de vital importancia pues, para que la democracia exista, se requiere de un efectivo sistema de comunicación política, en sentido amplio, que informe y que anime a las personas a participar, haciendo que se interesen por la política. Esto se vuelve especialmente importante a la medida en que las sociedades se vuelven más grandes y más complejas (McChesney p.13, 1997, citado por Pavlik, 2005).

En los últimos años, las organizaciones en favor de políticas de apertura y transparencia, de los denominados datos abiertos han ido en crecimiento (Sánchez, Bolaños, Magallón, & Anderica, 2015). Por ello, en épocas recientes se han tenido que desarrollar nuevos retos en los que se involucra a la sociedad. Se están generando nuevas formas de enfrentarse a los problemas de gobernanza local y global; entre

ellos destaca la colaboración y participación por parte de la ciudadanía a la hora de aportar soluciones a problemas comunes. Las estructuras clásicas de gestión pública deben adaptarse a los nuevos desafíos y aprovechar las oportunidades que la nueva economía colaborativa nos está empezando a plantear (Magallón, 2014).

2.4 Resumen del capítulo 2

El México contemporáneo requiere de una prensa que esté dirigida a la sociedad civil. Sin embargo, actualmente, el trabajo del periodista es considerado riesgoso, debido al incremento en los crímenes contra la libertad de expresión y el derecho a la información. Son pocos quienes dirigen sus esfuerzos hacia la sociedad por encima del poder público en su proceso informativo, ya que es este el aparente generador de responsables de tal vulnerabilidad, quienes prefieren limitar la información que el ciudadano mexicano promedio adquiere a través de medios como la televisión o la radio.

Por lo tanto, se puede inferir que aun cuando algunos de los profesionales del periodismo tienen los niveles de ética que les permiten el leal desempeño de su labor se ven obligados en algún punto a resguardar su seguridad.

Por otro lado, y como ya se ha mencionado, otro factor determinante es que el acceso al derecho a la información pública tiene en América

Latina una vida relativamente nueva. Sin embargo, se ha convertido en uno de los detonadores del derecho a la información y para México no es la excepción, un ejemplo de ello es la “Ley Federal de Transparencia y Acceso a la Información Pública Gubernamental” (LFTAIPG), que ha ido modificando la vida pública de México.

Es importante tener presente que más allá de combatir la corrupción, lograr la rendición de cuentas y transparencia, el acceso a la información pública es una herramienta para mejorar la calidad de vida de los individuos y de la sociedad, por lo que no debemos conformarnos con obtener información reducida o equivocada.

Es aquí donde interviene también la educación, pues las personas deben conocer la información a la que tienen derecho, así como lo que pueden hacer con tal información una vez que está en sus manos, y muy posiblemente será el impacto del P.D. un elemento que permitirá dimensionar ello.

Pese a que en México no existen muchos medios que hayan incorporado a su redacción un equipo para hacer P.D., los casos aislados que encontramos son muestra de que se están dando pequeños pasos hacia un ejercicio periodístico que trabaje con bases de datos abiertos, con grupos interdisciplinarios y en pro del P.D.

En el capítulo siguiente además de definir a los Datos Abiertos, se explica su relación con el periodismo y el nuevo rol que toman en las investigaciones. Asimismo, se contextualizan los Datos Abiertos en un país como México.

Datos Masivos y Datos Abiertos

Capítulo **3**

3.1 Datos Masivos

Históricamente la recolección de datos ha sido una tarea ardua, engorrosa e intensiva. Esta, parece ser la razón principal por la cual los datos masivos ofrecen en primera instancia, la percepción de poder acceder, manejar y dar contexto a enormes cantidades de datos. De esta manera, el uso de las computadoras y los algoritmos para administrar datos parecen ofrecer un ahorro de tiempo y esfuerzo considerables, así como un mejor rendimiento en este tipo de tareas (Boyd & Crawford, 2012).

Los datos masivos, son un fenómeno cultural, tecnológico y académico, en el cual convergen elementos tecnológicos, analíticos y especulativos. Los primeros, están relacionados con la capacidad de las computadoras y los algoritmos para: recolectar, analizar, vincular y comparar grandes conjuntos de datos. El segundo, consiste en la posibilidad de identificar, en estos conjuntos, patrones para la toma de decisiones económicas, sociales, técnicas o legales. Los terceros, se refieren a una creencia sobre la posibilidad de extraer de los grandes conjuntos de datos, nuevas formas de inteligencia y conocimiento (Boyd & Crawford, 2012).

De acuerdo, con la *International Data Corporation*, el fenómeno de los datos masivos se define como “un conjunto de herramientas tecnológicas capaces de extraer valor, de manera económica, de grandes volúmenes de datos a gran velocidad para su captura, descubrimiento o análisis”. Este

Unidad de medida	Potencia de 10	Número de bits
Gigabyte	10^9	1,000,000,000
Terabyte	10^{12}	1,000,000,000,000
Petabyte	10^{15}	1,000,000,000,000,000
Exabyte	10^{18}	1,000,000,000,000,000,000

Figura 12. Equivalencias en bits.

fenómeno tiene tres características principales: los datos, su análisis y la presentación de resultados (Gantz & Reinsel, 2012, pág. 9).

Sin embargo, en opinión de Sha Wei y Carrotti-Sha (2016), la definición de datos masivos es por naturaleza, demasiado vaga; principalmente porque implica la gestión y el procesamiento de cualquier colección de datos por parte de un individuo, lo cual, conlleva a dos obstáculos: la capacidad computacional para su procesamiento y los límites de la interpretación humana.

El fenómeno de los datos masivos es un enfoque novedoso en la comprensión de cantidades enormes de datos: estructurados, no estructurados y semi-estructurados. Esta clase de datos, requiere de una inversión considerable en términos de tiempo y dinero para ser procesados y analizados con herramientas convencionales, como las bases de datos relacionales (Barranco, 2012).

Cuando se estudia el tema, surge una disyuntiva acerca de la cantidad de datos ideal para que estos, sean considerados como masivos. Entre los expertos, se consideran como datos masivos a toda aquella cantidad de datos en un intervalo entre los *petabytes* y los *exabytes* (Barranco, 2012).

Es importante mencionar que, los datos masivos están definidos por el paradigma de las tres “V”, es decir, por el Volumen, la Variedad y la Velocidad de procesamiento y análisis necesario para interpretarlos (Barranco, 2012).

Cuando se trabaja con datos masivos, se requiere de contar con una taxonomía capaz de permitir navegar entre enormes cantidades de datos, de distinta naturaleza. Soares (2012) propone una breve clasificación de datos masivos a través de cinco elementos, como puede observarse en la figura 13.

En 2012, la *International Data Corporation* propuso, otra clasificación de posibles datos susceptibles de conformar un conjunto de datos masivos. En ella, se considera que los datos masivos se han vuelto visibles, gracias al mejoramiento en las técnicas de gestión, a través de la digitalización. Transformar información analógica hacia una representación numérica, permite recopilar y analizar información en cortos periodos de tiempo; además, otorga beneficios en su almacenamiento y procesamiento. Sin embargo, el surgimiento de la computadora, así como de la digitalización, no son razones suficientes para explicar el origen de los datos masivos. Para ello,

fue necesaria una transformación en el paradigma de organización y comprensión del mundo (Mayer-Shönberger & Cukier, 2013).

Para Milan y Gutiérrez (2015, pág. 12) el alcance del término “datos masivos”, se refiere a las diversas bases generadas por agencias gubernamentales, la extraordinaria cantidad de documentos en video y audio, texto, enlaces y etiquetas, resultado de la distribución y almacenamiento en línea. La información generada por la interacción humana en las plataformas sociales, los procesos de indexación de las actividades en internet y los metadatos resultantes de la vigilancia digital de los gobiernos.

Los datos masivos representan un episodio más en la historia de los mecanismos con los

Figura 13. Clasificación de datos masivos.

Medio	Datos
Web y Medios Sociales	Tuits, contenido web, datos de navegación, publicaciones de Facebook
Biometría	Reconocimiento facial y genética
Máquina a Máquina	Lecturas RFID, señales GPS
Datos Transaccionales	Registros de facturación, registros de telecomunicaciones
Generado por Humanos	Notas de voz, correo electrónico, documentos electrónicos

cuales, el ser humano ha buscado cuantificar y comprender su entorno (Mayer-Shönberger & Cukier, 2013). Esta nueva concepción de los datos, permite medir fenómenos ajenos a la realidad inmediata como: los latidos del corazón, la localización de una persona o las pautas con las que una persona realiza sus compras. El poder representar cualquier suceso de forma digital permite el fenómeno de la *datificación*.

La *datificación*, consiste en plasmar un fenómeno en un formato cuantificado para que pueda ser tabulado y analizado. Es importante mencionar que existe diferencia entre *digitalización* y *datificación*. La primera, se refiere a la transformación de información analógica en secuencias binarias de ceros y unos. Mientras que, la segunda, consiste en la representación y cuantificación de información por medio de datos factibles de ser medidos y analizados (Mayer-Shönberger & Cukier, 2013).

Según Boyd y Crawford (2012), los datos masivos no solo involucran grandes conjuntos de datos, herramientas y procedimientos para su análisis y manipulación; sino también, un cambio computacional en la manera de pensar e investigar. En este sentido, los datos masivos pueden considerarse como un sistema de conocimiento emergente, donde, las propiedades informativas de los propios objetos de conocimiento y las redes de comunicación, ofrecen la posibilidad de nuevas formas de entendimiento, a partir de la sistematización y análisis de grandes conjuntos de datos.

Lev Manovich (2011), ha identificado a tres

tipos de personajes involucrados en el fenómeno de los datos masivos: los creadores de datos, aquellos quienes tienen los medios para recolectarlos y finalmente, quienes cuentan con las habilidades para analizarlos. Boyd y Crawford (2012), han reflexionado sobre el rol de estos tres elementos, sobre todo en el empoderamiento de los profesionales, capaces de aprovechar las bondades de los datos, contra aquellos sin los conocimientos técnicos suficientes, quienes estarán del lado negativo de la nueva brecha digital: ricos en datos masivos contra los pobres en datos masivos.

Desde el periodismo, los datos masivos representan ideas emergentes respecto a bases de datos, algoritmos, métodos computacionales y procesos relativos al paradigma de la cuantificación. Aquí la parte masiva de los datos no se relaciona con el tratamiento masivo de datos, sino con la gran cantidad de posibilidades para el cambio social.

3.2 Definiendo los Datos Abiertos

Anteriormente se abordó de manera conceptual el fenómeno de los datos masivos y cómo estos influyen en la forma en la que concebimos el mundo actual. Ahora bien, es importante hacer énfasis que, en contados casos, los periodistas, en su labor se enfrentan con datos masivos.

Usualmente, trabajan con una derivación de ellos, conocidos como Datos Abiertos.

Los Datos Abiertos es una filosofía cuyo objetivo principal es poner a disposición de la sociedad aquellos datos generados por las administraciones públicas en formatos idóneos para ser manipulados. La idea central consiste en permitir a los ciudadanos analizar, utilizar y distribuir conjuntos de datos gubernamentales, con la finalidad de fortalecer las iniciativas de gobierno abierto e inteligente. El adjetivo abierto se refiere a que los formatos de los datos deben ser no propietarios; es decir, no pueden depender de una entidad o de una herramienta propietaria de una entidad. Como ejemplo de un formato abierto sería CSV (Comma-Separated Values) o XML (Extensible Markup Language), mientras que los formatos propietarios serían Word, Excel, etc., (Portal de Datos Abiertos de Barcelona, 2016).

Un ecosistema de Datos Abiertos debe ser capaz de proporcionar a los usuarios las siguientes tareas: descubrir la existencia de datos, acceder a ellos con fines de investigación y análisis, encontrar información detallada sobre el origen y descripción de los datos y la capacidad de reutilizar los instrumentos de recolección, compilación y agregado de datos, la comunidad efectiva con las agencias creadoras y distribuidoras de datos y compartir este conocimiento con otros usuarios (Open Data Foundation, 2016).

El sitio de Datos Abiertos del gobierno de la ciudad de Barcelona ha propuesto un decálogo de Datos Abiertos. Entre las características

idóneas de los Datos Abiertos se encuentran: públicos, detallados, actualizados, accesibles, automatizados, sin registro, abiertos y libres (Portal de Datos Abiertos de Barcelona, 2016).

En esencia los Datos Abiertos están definidos por el concepto de apertura. “Son datos que pueden ser utilizados, reutilizados y redistribuidos libremente por cualquier persona, y que se encuentran sujetos, cuando más, al requerimiento de atribución y de comportarse de la misma manera en que aparecen”. De acuerdo con Poikola (2009) son cuatro las características de los Datos Abiertos:

1. Disponibilidad y acceso,
2. Reutilización y redistribución,
3. Participación universal y
4. Ausencia de restricciones comerciales.

El término Datos Abiertos está relacionado con el de reutilización de información. Aunque los términos pueden parecer similares, el concepto Datos Abiertos persigue ofrecer los datos en formatos completamente libres (no propietarios) y no se contempla el pago por el uso de datos (deben ser gratuitos). Sin embargo, la reutilización de la información de las administraciones públicas prevé la posibilidad del pago por el uso de estos datos y la publicación de estos en cualquier formato (Portal de Datos Abiertos de Barcelona, 2016).

3.2.1 Clasificación

Existen diferentes clasificaciones de los Datos Abiertos. El portal de Datos Abiertos de la ciudad de Barcelona considera los formatos CSV, XML, RDF, RSS, ODATA como los principales de Datos Abiertos.

A continuación, se explica de manera general las características de los formatos antes mencionados:

■ **CSV** (Valores Separados por Coma) es un tipo de documento de texto plano para representar datos tabulados en columnas separadas por comas (o punto y coma) y filas separadas por saltos de línea. Las extensiones que utiliza son .csv o .txt. Es muy sencillo de utilizar y en muchos casos es trivial exportar los datos de una hoja de Excel a un formato CSV.

■ **XML** (Lenguaje de Etiquetado Extensible) es un metalenguaje simple que permite la interpretación de datos para diferentes lenguajes. Es el estándar para el intercambio de información estructurada entre diferentes plataformas. Muchas bases de datos permiten la exportación de sus datos en formato XML.

■ **RDF** (Infraestructura para la Descripción de Recursos) es un modelo universal basado en XML que permite intercambiar y enlazar a través de diferentes aplicaciones de datos

y recursos sin que pierdan su significado, lo que facilita la reutilización y enriquecimiento de los recursos de la web.

■ **RSS** (Sindicación Realmente Sencilla) es un formato basado en XML para la ordenación de contenidos de páginas web actualizados frecuentemente, como pueden ser las noticias, blogs o podcast. Aunque se usa mucho en la web no es el más aconsejado para Datos Abiertos debido a que la información que contiene no suele ser la más detallada.

■ **ODATA** (Open Data Protocol) es un protocolo abierto que permite realizar directamente consultas y selecciones sobre los conjuntos de datos ofrecidos, así como descargar los resultados de esas operaciones.

Béjar (2012) en un excelente trabajo de visualización muestra de manera didáctica la clasificación de Datos Abiertos en función de su nivel de apertura. En esta clasificación, propuesta por Tim Berners-Lee, compuesta por 5 niveles se asigna una estrella para determinar el nivel de apertura de un conjunto de datos. Parte de los formatos estáticos de presentación de datos (PDF) hacia los formatos de hojas de cálculo XLSX y CSV, para terminar en formatos de enlace y descripción de datos como son el RDF (Resource Description Framework) y LOD (Linked Open Data).

Figura 14. Elaboración propia basada en la clasificación de Tim Berners-Lee. (Fuente: <http://5stardata.info/es/> Consultado el: 15 de octubre de 2017)

Nivel	Descripción
★	Publica datos en la Web (con cualquier formato) y bajo una licencia abierta.
★ ★	Publicar datos estructurados (ej.: Excel en vez de una imagen de una tabla escaneada).
★ ★ ★	Usa formatos no propietarios (ej.: CSV en vez de Excel).
★ ★ ★ ★	Usa URLs para denotar cosas, así la gente puede apuntar a estas.
★ ★ ★ ★ ★	Enlaza datos a otros datos para proveer contexto.

3.3 Datos Abiertos en México

En materia de Datos Abiertos en México, en febrero de 2015 el presidente de la República, Enrique Peña Nieto publicó un decreto para regular los Datos Abiertos. Dicho decreto tiene por “objeto regular la forma mediante la cual, los datos de carácter público generados por las dependencias y entidades de la Administración Pública Federal y por las empresas productivas del Estado, se pondrán a disposición de la población como Datos Abiertos, con el propósito de facilitar su acceso, uso, reutilización y redistribución para cualquiera que sea

el fin, conforme a los ordenamientos jurídicos aplicables” (Diario Oficial de la Federación).

El decreto sugiere la conformación de un catálogo de Datos Abiertos integrado por los conjuntos de datos de las dependencias y entidades, así como las empresas productivas del Estado descargables en formatos abiertos, en un portal de internet. Así mismo establece las características deseables de un conjunto de Datos Abiertos, entre ellas se mencionan: su carácter gratuito, la no discriminación, el libre uso, su legibilidad por máquina, su integridad, su calidad como fuente primaria, así como su permanencia y su carácter oportuno (Diario Oficial de la Federación). En el año 2015, como resultado del decreto presidencial, surge la

plataforma <http://www.datos.gob.mx>, sitio oficial de Datos Abiertos del Gobierno de la República. Asegura ser el lugar de consulta y descarga en formato libre de todos los Datos Abiertos que posee el Gobierno de la República, así como de estados, municipios y órganos autónomos que han decidido sumarse a la iniciativa para poner a disposición de la ciudadanía información pública. Sin embargo, con base en el informe sobre Datos Abiertos en México de la Organización para la Cooperación y el Desarrollo Económico (OECD) en estos temas aún queda mucho camino por recorrer, por ejemplo, un esfuerzo sistemático para producir un gran impacto en la economía y la sociedad mexicanas por medio de la aceleración de la economía digital y de la mejora de la actividad pública.

La Organización para la Cooperación y el Desarrollo Económico (OECD) publicó a mediados de 2016 un estudio en materia de Datos Abiertos en México donde se sugieren algunas propuestas para fortalecer las políticas de Datos Abiertos en nuestro país. De acuerdo con el informe ejecutivo para los Datos Abiertos en México, los Datos Abiertos Gubernamentales (DAG) aún no generan el impacto deseado y sostenible en la actividad económica y la sociedad mexicana. (OECD, 2016)

El estudio analiza los esfuerzos que, hasta la fecha, el gobierno federal ha realizado para establecer un marco para la gobernanza y la política sobre gobierno abierto y Datos Abiertos gubernamentales, examina la implementación

de iniciativas tanto a nivel federal como a nivel local. Así mismo destaca la existencia de oportunidades para aprovechar plenamente los beneficios de los DAG en México. (OECD, 2016)

Como sugiere el informe de la OECD no basta con que el gobierno mexicano libere bases de Datos Abiertos, sino que resulta fundamental involucrar a diferentes actores en el uso de los mismos. De nada sirve que el gobierno coloque información si nadie pone en duda su veracidad y le da un uso ya sea en la toma de decisiones o bien para exigir al gobierno mayor transparencia.

3.3.1 El portal de Datos Abiertos de la República Mexicana

El portal de Datos Abiertos forma parte de un movimiento a nivel global encargado de trascender las iniciativas de transparencia y rendición de cuentas. Su finalidad es fomentar el uso de datos e información pública en el diseño e implementación de políticas públicas, así como generar espacios de debate y participación ciudadana.

[Datos.gob.mx](http://datos.gob.mx)¹⁶ es la plataforma digital oficial del gobierno de la República mexicana, en ella, se albergan bases de datos abiertas cuyas características técnicas y legales permiten usar, reusar y distribuir la información en cualquier lugar o momento. El sitio se compone de tres secciones: datos, blog y herramientas.

Su función principal es permitir la consulta y

¹⁶ www.datos.gob.mx
(Consultado el: 1 de octubre de 2017)

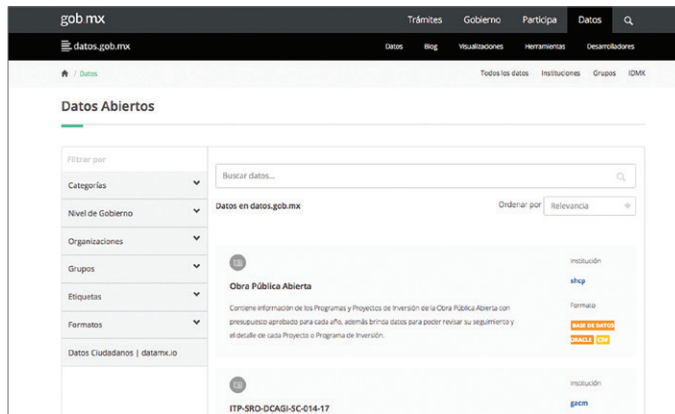


Figura 15. Página de la sección de datos de la plataforma digital del gobierno de la República Mexicana. (Fuente: <https://datos.gob.mx/busca/dataset> Consultado el: 15 de octubre de 2017)

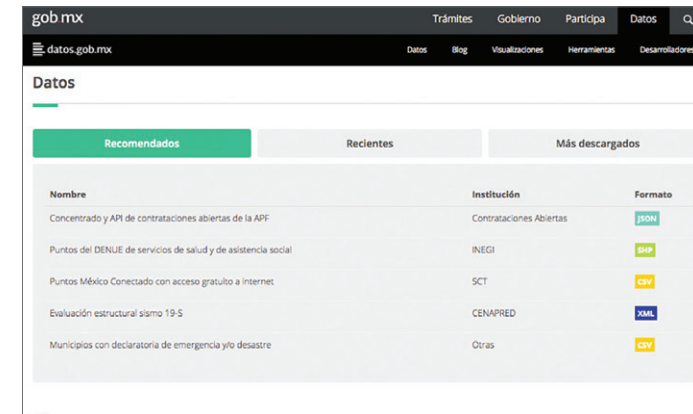


Figura 16. Ejemplo de la exploración de datos de la plataforma. Disponible en: <https://datos.gob.mx>

descarga de datos sobre las actividades del gobierno federal, estados, municipios y órganos autónomos en formato libre.

La sección de *Datos* permite acceder a las bases de datos a partir de su nombre, dependencia productora y tipo de formato. Sin embargo, la forma más sencilla de navegar en el portal es mediante la exploración de las bases “recientes”, las “recomendadas” y las “más descargadas” (Figura 15). Este menú sencillo dota de interactividad al sitio, pues permite al usuario encontrar de manera rápida los datos que busca.

La sección del blog, funciona como medio para dar a conocer eventos y casos de uso, donde se hace referencia a la aplicación de Datos Abiertos tanto en México como en el mundo.

Asimismo, la plataforma cuenta con una sección de herramientas (Figura 17), en ella se encuentran albergados recursos como: mapas, aplicaciones web para el uso, visualización y

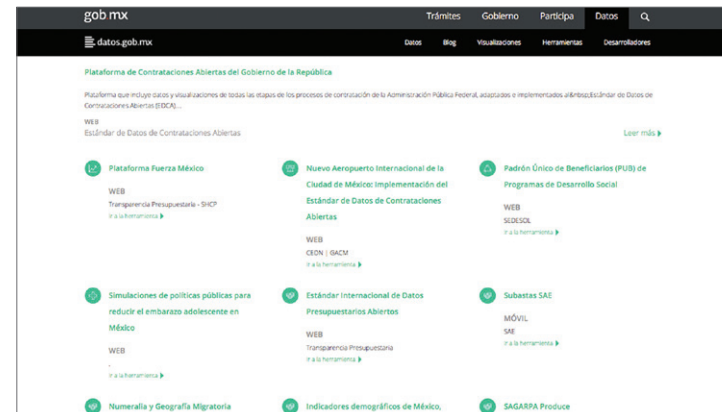


Figura 17. Ejemplo del catálogo de herramientas. Disponible en: <https://datos.gob.mx/herramientas>

generación de conocimiento a partir de D.A.

Las diferentes herramientas que en dicha sección se muestran, buscan ayudar a resolver problemas en materia de: cultura y turismo, desarrollo sostenible, economía, educación, energía y medio ambiente, finanzas y contrataciones, geoespacial, gobiernos locales, infraestructura, salud y seguridad, así como justicia.

En la Tabla 1 se muestran las 40 herramientas desarrolladas por las diferentes dependencias de la administración pública federal, por nombre, dependencia de origen y tipo de herramienta.

3.4 Resumen del capítulo 3

El incremento en el uso de la computadora y los algoritmos para la recolección, creación, tratamiento y difusión de grandes colecciones de datos permiten un ahorro de tiempo y esfuerzo considerables, así como un mejor rendimiento en este tipo de tareas. Estas bondades han configurado el paradigma de los datos masivos, como fenómeno cultural, tecnológico y académico, en el cual convergen elementos tecnológicos, analíticos y especulativos.

El término datos masivos, hace referencia a diversas bases de datos generadas por agencias gubernamentales, en la forma de: documentos en video y audio, texto, enlaces y etiquetas, resultado de la distribución y almacenamiento en línea. La información generada por la interacción humana en las plataformas sociales, los procesos de indexación de las actividades en internet y los metadatos resultantes de la vigilancia digital de los gobiernos.

El movimiento de Datos Abiertos es una iniciativa que tiene carácter internacional impulsada

Tabla 1. Ejemplo del catálogo de herramientas. Disponible en <https://datos.gob.mx/herramientas>

No.	Nombre	Dependencia	Descripción
1	Atlas Turístico de México	Secretaría de Turismo	Móvil
2	Subastas SAE	SAE	Móvil
3	Numeralia y Geografía Migratoria	CONAPO	Web
4	Indicadores Demográficos de México, periodo 1990 - 2030	CONAPO	Web
5	Sagarpa Produce	SAGARPA	Móvil
6	Atlas de Complejidad Económica de México	Secretaría de Hacienda y Crédito Público	Web
7	Prodecon	PRODECON	Móvil
8	Herramienta de Datos para el Cambio Climático	Coordinación de Estrategia Digital Nacional	Web
9	RadarCiSalud	Secretaría de Salud	Móvil
10	Cálculo consumo Hogar	Secretaría de Energía	Web
11	Regionalización Funcional de México	SEDATU	Web
12	Contrataciones Abiertas en México	Coordinación de Estrategia Digital Nacional	Web
13	Herramienta de Datos para el Desarrollo Sostenible	Coordinación de Estrategia Digital Nacional	Web
14	Portal Interactivo Hidrometeorológico	Servicio Meteorológico Nacional	Web
15	IMSS Digital	Instituto Mexicano del Seguro Social	Móvil
16	UbicaPue	Secretaría de Finanzas y Administración del Estado de Puebla	Móvil
17	Veracruz Ciudad	Municipio de Veracruz	Móvil
18	Reconstrucción MX	Coordinación de Estrategia Digital Nacional	Web
19	CONEVAL	Consejo Nacional de Evaluación de la Política de Desarrollo Social	Móvil
20	Mapa Digital de México	Instituto Nacional de Estadística y Geografía	Web

21	México en Cifras	Instituto Nacional de estadística y Geografía, INEGI	Móvil
22	Directorio Estadístico Nacional de Unidades Económicas (DENUE)	Instituto Nacional de estadística y Geografía, INEGI	Móvil
23	Entidades Federativas	Secretaría de Hacienda y Crédito Público	Web
24	Obra Pública Abierta	Secretaría de Hacienda y Crédito Público	Web
25	Atlas Municipal de Personal de Educación Básica y Especial - maestros (2013)	Coordinación de Estrategia Digital Nacional	Web
26	Atlas Municipal de Alumnos de Educación Básica y Especial (2013)	Coordinación de Estrategia Digital Nacional	Web
27	Consulta de la Cartografía Minera Digitalizada	Secretaría de Economía, Dirección General de Regularización Minera	Web
28	Consultas Públicas en el proceso de evaluación de impacto ambiental	Secretaría de Medio Ambiente y Recursos Naturales	Web
29	Guía Pemex	Petróleos Mexicanos	Móvil
30	SAT Móvil	Servicio de Administración Tributaria	Móvil
31	Mappir	SCT (Unidad de Tecnologías de Información y Comunicación)	Móvil
32	GeoInfoMex	Servicio Geológico Mexicano	Web
33	Obra Pública e Infraestructura SCT	Secretaría de Comunicaciones y Transportes	Web
34	Registro Nacional de Datos de Personas Extraviadas o Desaparecidas RNPED	Secretario Ejecutivo del Sistema Nacional de Seguridad Pública	Web
35	Asegurados y salarios en el IMSS	Coordinación de Estrategia Digital Nacional	Web
36	Mortalidad Materna en México	Coordinación de Estrategia Digital Nacional	Web
37	Pobreza y rezago social en México	Coordinación de Estrategia Digital Nacional	Web
38	Utilización del fondo para la atención de emergencias y reconstrucción en desastres naturales en 2013	Coordinación de Estrategia Digital Nacional	Web
39	Incidencia Delictiva del Fuero Federal 2012 - 2015	Coordinación de Estrategia Digital Nacional	Web
40	Inundaciones y proyectos de reconstrucción en Tampico, Tamaulipas durante el huracán Ingrid	Coordinación de Estrategia Digital Nacional	Web

por los países desarrollados en conjunto con la Organización para la Cooperación y el Desarrollo Económico (OECD). Su principal objetivo es fomentar las políticas públicas en materia de acceso a la información, transparencia y rendición de cuentas. En México, desde el año 2015 la presidencia de la República fomentó un decreto para fortalecer el estado de los Datos Abiertos gubernamentales, por medio de la implementación de un sitio web que proporciona diversos servicios entorno a los datos.

Los Datos Abiertos es una filosofía cuyo objetivo principal es poner a disposición de la sociedad aquellos datos generados por las administraciones públicas en formatos idóneos para ser manipulados. Tim Berners-Lee ha propuesto una clasificación de ellos basados en 5 niveles:

■ **Nivel 1:** Publicar datos en la web en cualquier formato con una licencia abierta.

■ **Nivel 2:** Publicar datos estructurados en la web como XLSX.

■ **Nivel 3:** Publicar datos en formatos no propietarios como CSV.

■ **Nivel 4:** Publicar datos en formato de Resource Description Framework (RDF).

■ **Nivel 5:** Enlazar datos con otros a través de Linked Open Data (LOD).

Este sitio se llama Datos.gob.mx, se encuentra en esa misma dirección, la cual es la plataforma digital oficial del gobierno de la república mexicana. Aquí se albergan bases de datos abiertas cuyas características técnicas y legales permiten usar, reusar y distribuir la información albergada en ella en cualquier lugar o momento. El sitio se compone de tres secciones: datos, blog y herramientas. Su principal objetivo es permitir la consulta y descarga de datos sobre las actividades del gobierno federal, estados, municipios y órganos autónomos en formato libre.

La iniciativa de gobierno abierto también proporciona herramientas para el uso de las bases de datos a través de dos mecanismos: aplicación móvil o web. Con un total de 40 opciones los usuarios del portal de datos abiertos pueden explorar algunos conjuntos de datos.

En el siguiente capítulo se presentan las herramientas más comúnmente utilizadas por los periodistas a la hora de aplicar las fases de la metodología del P.D. Además se describen algunos de los cursos disponibles en línea que, si bien no están enfocados al P.D., si tocan temáticas relacionadas con dicha especialización.

Herramientas de apoyo para el Periodista de Datos

Capítulo **4**

4.1 Análisis de herramientas de apoyo para el Periodista de Datos

La era tecnológica en la que vivimos, permite a quien tenga los recursos y el conocimiento desarrollar material noticioso dentro de la red. Es por ello que el periodista está obligado a mantenerse en constante actualización sobre contenidos digitales, nuevos software, tendencias, etc.

El periodista del siglo XXI debe aprender a manipular grandes cantidades de información, mismas que serán la base de su investigación. En la actualidad, la labor del periodista se define por las herramientas que utiliza, tanto en la búsqueda u obtención de datos, como en la limpieza, análisis y visualización. Se debe tener el criterio necesario para la elección de las mismas, además de ciertos conocimientos específicos de su profesión. A la hora de elegir las herramientas a utilizar, también se debe tener en cuenta que muchas de ellas solo están disponibles en inglés, por lo que el dominio de dicho idioma es sin lugar a dudas necesario.

Hoy en día, se pueden obtener datos por medio de buscadores como Google, a partir de una búsqueda específica. Hay diversas formas para su recuperación, por lo que el periodista debe estar familiarizado en la búsqueda de cierto tipo de información especializada.

Cada día se abren más portales en la web para

la obtención de Datos Abiertos y centros de datos en la red, que sirven como motor de búsqueda para la recopilación de los mismos. Ya sean de gobiernos o de organizaciones no lucrativas, existe una mayor apertura de datos para el periodista. Esta apertura de datos es significativa, pues da pie a que se generen cada vez más artículos basados en el análisis de bases de datos.

Lo que no hagan los gobiernos en materia de Datos Abiertos lo hará la comunidad, las organizaciones de la sociedad civil y los medios de comunicación (Crucianelli, 2013). Es por ello que cada vez hay más sitios enfocados en compartir datos como *La Nación Costa Rica*¹⁷. Así mismo existen otros sitios donde es posible obtener datos de investigación como *UK Data Archive*¹⁸ en el cual se puede tener acceso a información de manera gratuita, aunque existe otra parte del sitio en donde solo se puede acceder mediante suscripción.

4.1.1 Herramientas de Búsqueda de datos

Cuando se lleva a cabo el proceso de obtención de datos, es necesario que la búsqueda sea muy específica, o bien, manejar ciertas herramientas pensadas exclusivamente en ayudar a realizar la búsqueda. Esta parte del proceso es fundamental, pues en el P.D. es imprescindible contar con

¹⁷ <http://www.nacion.com/gnfactory/investigacion/periodismo-datos> (Consultado el: 19 de octubre de 2017)

¹⁸ <http://www.data-archive.ac.uk> (Consultado el: 11 de noviembre de 2017)

ellas para poder empezar a plantear una hipótesis periodística.

Para iniciar la búsqueda de datos Google es de gran ayuda, pues mediante el uso de los filtros de geolocalización y la selección del tipo de archivo a encontrar, el periodista puede darse una idea de las bases de datos disponibles en la red para identificar cuáles resultarían relevantes para su investigación.

En el Manual de Periodismo de Datos (2012) se pueden encontrar algunas herramientas web a considerar a la hora de realizar búsquedas de datos, mismas que ayudan a indagar en la obtención de determinados datos. Sitios como *Whois*¹⁹, donde el usuario realiza una búsqueda sobre algún sitio o URL, para obtener información básica del mismo como IP, servidor, fecha de creación, capturas de pantalla del sitio a lo largo de la historia, etc.

También, existen otro tipo de herramientas pensadas en ayudar a comprender la forma en la que se comparte la información, algunas de estas herramientas son:

■ **YouTube**²⁰: página de Internet en la cual los usuarios pueden compartir y subir videos. Dentro de las métricas del sitio, los usuarios pueden conocer el número de veces que el video ha sido reproducido o bien, identificar los más vistos.

Otro tipo de herramientas que permiten al usuario conocer las tendencias en redes o en la web,

sobre el interés de los usuarios en determinados sitios son:

■ **Storyful MultiSearch**²¹: herramienta de código abierto que busca en múltiples plataformas de redes sociales sobre algún tema particular.

■ **Google Trends**²²: arroja información sobre las tendencias de búsqueda en Google. Permite conocer los temas más buscados por día, hora o año, así como comparar los intereses por regiones y tiempo. También, ayuda a ubicar los lugares específicos de donde provienen determinadas búsquedas.

■ **Twitter**²³: es un microblogging que ayuda a medir tendencias por medio de sus hashtags, con la suma de estos se crean Trending Topics en cada país y esto ayuda a medir las tendencias en esta red social.

■ **Social Mention**²⁴: es una herramienta que ayuda a monitorear varios sitios de redes sociales a la vez, analiza la información de forma profunda y mide la influencia de temas por tópicos.

Las redes sociales le han dado otro sentido a la noticia, antes los grandes medios determinaban el valor de cada nota y el usuario no tenía posibilidad de interactuar con los periodistas o dar su opinión sobre ciertas notas. Ahora la

¹⁹ <http://whois.domaintools.com> (Consultado el: 3 de noviembre de 2017)

²⁰ <https://www.youtube.com> (Consultado el: 20 de octubre de 2017)

²¹ <https://chrome.google.com/webstore/detail/storyful-multi-search/hkglibabhnbjmaccpajjakojeacnaf> (Consultado el: 10 de octubre de 2017)

²² <https://trends.google.com/trends> (Consultado el: 18 de noviembre de 2017)

²³ <https://twitter.com> (Consultado el: 18 de noviembre de 2017)

²⁴ <http://socialmention.com> (Consultado el: 29 de octubre de 2017)

situación es totalmente diferente, la opinión de los lectores es básica si los medios quieren ser escuchados; es por ello que ahora el periodista tiene la labor de medir la tendencia de los temas más discutidos, además, debe conocer de lo que se habla y comparte en las redes. Como Tascar Lara (2009) la define en su texto “La nueva esfera pública. Los medios de comunicación como redes sociales”, la opinión de los lectores es básica si se quiere que los medios se reconecten con la sociedad y se acabe la crisis de autoridad. Por ello, los medios necesitan crear espacios para que los lectores interactúen con el medio y se sientan parte de él. Las herramientas que ayudan al periodista a obtener datos de la red son de gran ayuda, pues le permiten explorar y conseguir información específica de los temas sobre los que giran sus investigaciones. Conocer este tipo de herramientas y saber utilizarlas representaran una disminución en el tiempo invertido explorando la red en busca de bases de datos que sustenten sus investigaciones; de lo contrario el proceso de obtención podría ser largo y confuso, pues no toda la información se encuentra en formatos que permitan su fácil procesamiento.

Un punto importante a destacar es que, en nuestro país no toda la información se encuentra disponible en formato digital, por lo que en muchas ocasiones será necesario realizar solicitudes vía transparencia. Para ello se deberá hacer uso de otro tipo de herramientas que permitan convertir documentos de un formato no editable a otro procesable para facilitar su análisis.

4.1.2 Herramientas de Scraping

Existen varios términos utilizados al hablar de obtención de datos, entre ellos tenemos el del *scraping*. De acuerdo con Rigoberto Carvajal, Ingeniero en Computación del Instituto Tecnológico de Costa Rica, se conoce como *scraping* al arte de controlar de manera automatizada sitios web con el propósito de extraer datos publicados que son de nuestro interés (Manual de Periodismo de Datos Iberoamericano., 2015). Es decir, cuando la información contenida en un sitio web no está disponible para descargar, es momento de aplicar esta técnica, pues su propósito es traducir información contenida en Internet a formatos editables, con los que se es posible trabajar.

Por ello, resulta importante que, si el periodista se ve en la necesidad de *scrapear* un sitio, cuente con herramientas especializadas para poder hacerlo. A continuación, se mencionan algunas de ellas:

■ **HTTrack²⁵**: Permite al usuario descargar un sitio de Internet a un directorio local, generando directorios reutilizables mediante la obtención del código HTML, imágenes y otros archivos obtenidos del servidor.

■ **JAVA, PHP y C#** (Lenguajes de programación): Permiten extraer información de sitios web mediante el uso de algunas líneas de código.

²⁵ <https://www.httrack.com>
(Consultado el: 4 de noviembre de 2017)

■ **Import.io**²⁶: captura información de sitios web con solo ingresar el *link* de un sitio y la información la recupera poniéndola dentro de una base de datos.

En la red, hay información que no se puede procesar porque de origen es un archivo de imagen o un archivo *Portable Document Format* (PDF), por ello el periodista de datos, tiene que hacer uso de software que sirva para extraerlos.

En el caso de México, algunas fuentes gubernamentales proporcionan información en los archivos como los que se mencionaron anteriormente, o incluso hacen entrega de datos por medio de fotocopias. Haciendo que el trabajo del periodista se convierta en elaborar bases de datos de manera casi artesanal.

Para agilizar este proceso existen herramientas como *Zamzar*²⁷, que es un convertidor gratuito que permite pasar un archivo en un formato definido a otro lo que permite al usuario trabajar con él. Para poder usarlo, se debe indicar la URL de la cual se desea convertir la información del formato de origen (usualmente JPEG o PDF) a un formato editable (Documento de Word o Excel, entre otros), una vez que se ha ingresado la URL y el formato final, el usuario deberá ingresar el correo electrónico donde desea recibir el archivo descargable de la información solicitada.

Existen herramientas que ayudan a obtener información de tablas y gráficos que pueden estar en formato de imagen, algunas de ellas son:

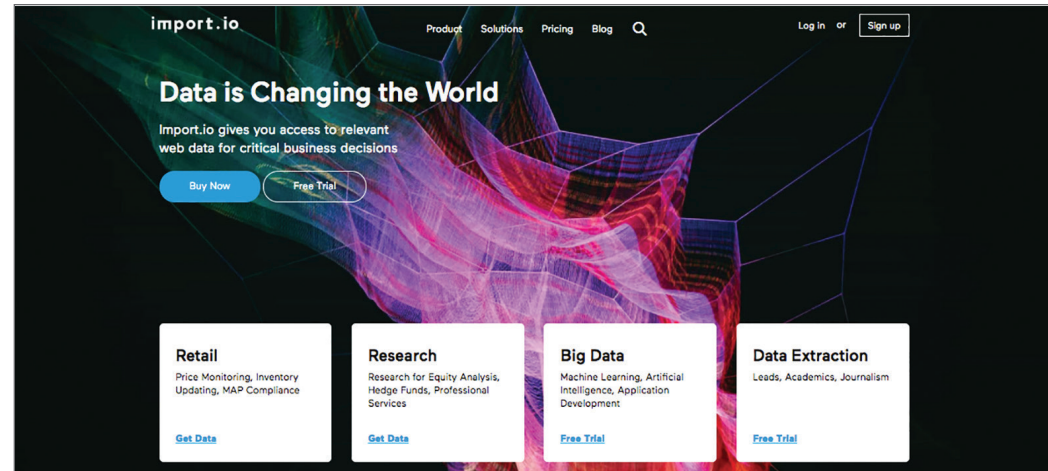


Figura 18. Sitio herramienta import.io (Fuente: <https://www.import.io/> Consultado el: 28 de octubre de 2017)

■ **Free OCR**²⁸: herramienta gratuita para realizar reconocimiento óptico de caracteres (OCR), le permite extraer texto de cualquier imagen.

■ **CometDocs**²⁹: convierte archivos en formato PDF y lo convierte a otro formato (Word, Excel, PowerPoint, imágenes, entre otros).

■ **Tabula**³⁰: convierte archivos que de origen están en formato PDF con el propósito de extraer la data que contienen en hojas de Excel o bien, en un archivo CSV (Comma-Separated Values).

El uso de este tipo de herramientas permitirá al periodista obtener información pertinente para su investigación, que bajo el esquema

²⁶ https://www.import.io (Consultado el: 28 de octubre de 2017)

²⁷ <http://www.zamzar.com> (Consultado el: 14 de octubre de 2017)

²⁸ <http://www.free-ocr.com> (Consultado el: 19 de octubre de 2017)

²⁹ <http://www.cometdocs.com> (Consultado el: 18 de octubre de 2017)

³⁰ <http://tabula.technology> (Consultado el: 6 de noviembre de 2017)

tradicional (recapturar la información) podría tomar demasiado tiempo y requerir de una mayor cantidad de recursos. Sin embargo, para poder hacer uso de ellas, es necesario que el periodista cuente con conocimientos básicos de programación sobre cómo las herramientas logran convertir información en formatos no editables a formatos como .XLS o .CSV que permiten su fácil procesamiento.

Contar con bases de datos en formatos editables, permitirá hacer uso de las herramientas auxiliares diseñadas para las fases subsecuentes de la metodología.

4.1.3 Herramientas de Limpieza de Datos

Muchas de las bases de datos que se recuperan mediante el *scraping*, tendrán espacios en blanco, comillas de más o bien, pueden no estar ordenadas. En esta parte del proceso de trabajo del periodista de datos, se tendrá que llevar a cabo una minuciosa limpieza de las mismas, para que sea posible establecer una jerarquía dentro de los datos obtenidos. Esta parte del proceso es fundamental y puede representar hasta un 90% del trabajo.

El periódico “*The Guardian*” en su manual de P.D., menciona algunas de las herramientas disponibles para ayudar al periodista en esta etapa, entre ellas podemos encontrar:

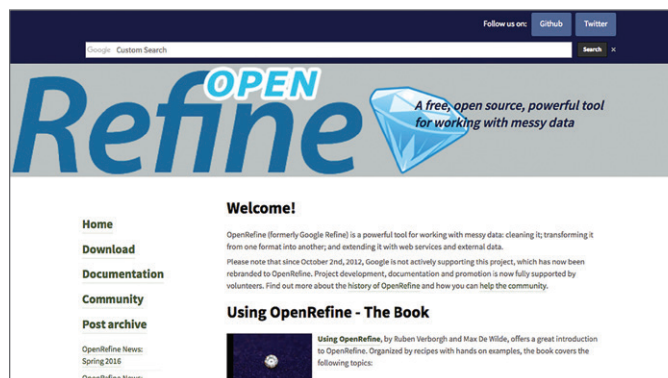


Figura 19. Sitio herramienta Open Refine (Fuente: <http://openrefine.org/> Consultado el: 21 de octubre de 2017)

■ **Open Refine³¹**: permite identificar errores en la base de datos para su corrección. Funciona mediante la implementación de algoritmos de agrupación, que identifican caracteres similares, agrupándolos para así, hacer más sencilla su corrección.

■ **Excel³²**: es un programa diseñado para la creación, modificación y manejo de hojas de cálculo, tiene la opción de limpiar una base de datos por medio de la utilización de filtros y la aplicación de fórmulas.

Después de analizar las diferentes herramientas diseñadas para esta fase de la metodología del P.D., pudimos notar que existen realmente muy pocas que ayuden a sistematizar este proceso, debido en buena medida porque no existe un formato estándar de bases de datos. Cada una cuenta con particularidades específicas que deben ser consideradas a la hora de limpiarlas (estandarizarlas).

³¹ <http://openrefine.org> (Consultado el: 10 de noviembre de 2017)

³² <https://products.office.com/en-us/excel> (Consultado el: 20 de octubre de 2017)

Si bien, las herramientas antes mencionadas ayudan a que el proceso de estandarización sea mucho más sencillo, la realidad es que es una fase que requiere de la inversión de tiempo y de conocimiento de los elementos que la confirman. De la correcta estandarización de la base de datos a trabajar, dependerá que el análisis de los datos sea realizado apropiadamente.

4.1.4 Herramientas para la Exploración o Análisis de Datos

Una vez que el periodista cuenta con una base de datos ordenada y sistematizada, el siguiente paso es el del análisis de la información. Aquí se debe tener en cuenta que los números por sí solos no dicen nada, aquí los conocimientos del periodista son fundamentales, ya que con ellos será capaz de cotejar la información cuantitativa obtenida, con el contexto del fenómeno a indagar.

En esta parte de la metodología, se recomienda al periodista acercarse a un experto en análisis de datos, con la intención de que este pueda explicarle la metodología con las que trabajan las herramientas diseñadas para el análisis de datos para llevarlo a cabo de una forma más eficiente.

Tomando como punto de partida, la hipótesis que se planteó al inicio de su investigación, el periodista será capaz de detectar puntos

relevantes dentro de su base de datos. Si el periodista cuenta con conocimientos en matemáticas, informática y estadística, tiene mayores posibilidades de descubrir primicias entre los números, de lo contrario siempre puede hacer uso de las herramientas disponibles que pueden ayudar a que esta parte del proceso sea mucho más sencilla, algunas de las más utilizadas son:

■ **R³³**: lenguaje y entorno de programación para el análisis estadístico.

■ **Tableau³⁴**: herramienta que permite analizar y visualizar información. Mediante el uso de tableros interactivos, el usuario puede ir descubriendo historias detrás de los datos.

Para el análisis de datos existen tanto herramientas altamente sofisticadas, pensadas específicamente para el análisis de datos masivos como herramientas más sencillas, pensadas para periodistas.

Para esta fase de la metodología, es importante que el periodista tenga en mente que no siempre será necesario contar con conocimientos avanzados en programación. En muchas ocasiones, la aplicación de fórmulas básicas de estadística, serán suficientes para analizar sus bases de datos.

Para aquellas bases de datos más complejas de abordar, será necesario que se apoye de un programador o analista de datos, el cual puede ayudarle a abordarlas de forma mucho más sencilla.

33 <https://www.r-project.org>
(Consultado el: 16 de octubre de 2017)

34 <https://www.tableau.com>
(Consultado el: 20 de octubre de 2017)

4.1.5 Herramientas para la Visualización de Datos

Los datos obtenidos por el periodista son una serie de números, que mediante la visualización serán capaces de adquirir sentido, es decir, ser mucho más claros para el lector. De acuerdo a la historia a contar se deberán seleccionar los datos pertinentes para su representación visual.

Al visualizar la información se pueden encontrar otros patrones o conexiones interesantes que no se habían revelado durante la búsqueda de datos. Sin lugar a dudas, es un punto fundamental que puede ayudar al periodista a detectar nuevos patrones que permitan enriquecer o ampliar la nota periodística.

La elección del gráfico dependerá del artículo, por lo que es importante tener claro cuáles son los puntos importantes a visualizar ya que la gama de formas gráficas para una representación es muy amplia.

Es importante destacar que no todas las investigaciones basadas en datos deben estar acompañadas de una representación visual, en ocasiones habrá historias donde la investigación y los datos que la sustentan sean lo suficientemente claros como para permitirle al lector entender el artículo en su totalidad. La visualización es un acompañamiento que permitirá entender los datos, pero no un mandatorio.

Dentro de la variedad de software para visualización de datos encontramos desde una plantilla de cálculo como *Google Docs* hasta software



Figura 20. Sitio herramienta d3js (Fuente: <https://d3js.org/> Consultado el: 21 de octubre de 2017)

para crear gráficas como *Tableau Public*. Cada una de las diferentes herramientas contiene un tutorial que sirve como guía para que el periodista pueda crear diferentes visualizaciones.

Aquí incluimos un listado de algunas herramientas de visualización:

Herramientas generales o de diversos usos:

■ **Processing³⁵**: lenguaje de programación integrado de código abierto basado en Java, que permite la realización de visualizaciones interactivas multimedia.

■ **D3.js³⁶**: librería de JavaScript para la generación de visualizaciones complejas, pueden ser dinámicas e interactivas a partir de datos.

³⁵ <https://processing.org> (Consultado el: 20 de octubre de 2017)

³⁶ <https://d3js.org> (Consultado el: 11 de octubre de 2017)

Herramientas pensadas específicamente en el mapeo:

■ **CartoDB³⁷**: servicio Web para almacenar, visualizar y compartir mapas interactivos a través de bases de datos geoespaciales.

■ **Visualize Free³⁸**: mediante la combinación de visualización y preparación de datos, permite al usuario comenzar a visualizar los datos con los que está trabajando.

■ **My Maps³⁹**: herramienta desarrollada por Google que permite generar de forma sencilla mapas interactivos. Sustentada por Google Maps, facilita al usuario agregar ubicaciones, texto, fotos o video. El usuario solo necesita ingresar una base de datos con información sobre la ubicación para que la herramienta pueda detectarlos y convertir esa información en mapas.

En el caso de la visualización existe una amplia oferta de herramientas en línea que permiten realizar desde los más básicos gráficos (barras, pie, histogramas, etc.), hasta los más sofisticados mapas.

Pero, para que el periodista pueda sacar el mayor provecho de las mismas, será de gran utilidad que posea conocimientos básicos de diseño como forma, color, jerarquías, entre otros. Esto le permitirá seleccionar la mejor representación gráfica que acompañe a su investigación.

4.1.6 Herramientas más usadas por los expertos

Dentro del manual de Periodismo de Datos, periodistas del diario “The Guardian” hacen mención de algunas herramientas por las que tienen cierta afinidad a la hora de visualizar sus reportajes. Entre ellas podemos encontrar herramientas de Google como *Google Fusion Tables* o *Google Spreadsheets*, pues ellas permiten a los lectores interactuar con la información.

Dentro del mismo manual, podemos encontrar en el mismo apartado que periódicos como “La Nación” de Argentina hacen uso de diversas herramientas como *Excel* o *Google Spreadsheets* para edición, y software como *Junar* para compartir sus datos.

Algunas de estas herramientas pueden ayudar al periodista a realizar su trabajo de manera efectiva, sin la necesidad de convertirse en un programador.

Definir qué herramientas utilizar dependerá de la afinidad que el periodista tenga con ellas, pues se tiene una gran variedad de opciones. Por un lado, existen herramientas que son muy amigables y llevan de la mano al usuario y por el otro, también hay algunas mucho más complejas, que para su uso se requieren de conocimientos avanzados ya sea en matemáticas o bien en programación.

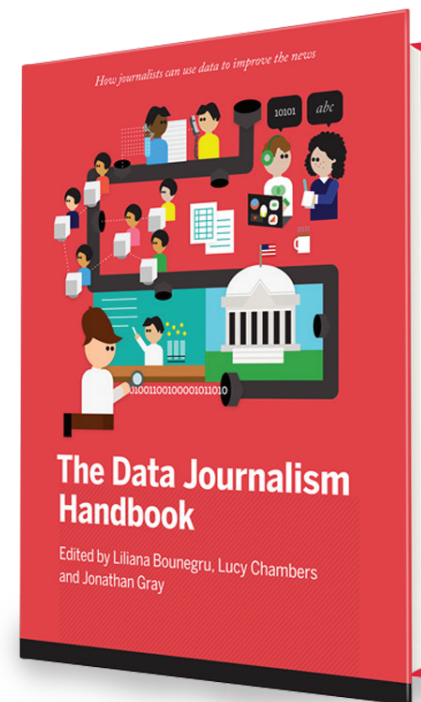


Figura 21. Portada del Manual de Periodismo de Datos. (Fuente: <http://interactivos.lanacion.com.ar/manual-data/> Consultado el: 4 de octubre de 2017)

³⁷ <https://carto.com> (Consultado el: 12 de octubre de 2017)

³⁸ <https://www.visualizefree.com> (Consultado el: 6 de noviembre de 2017)

³⁹ <https://www.google.com/maps/about/mymaps/> (Consultado el: 4 de noviembre de 2017)

En la siguiente tabla se muestra a modo de resumen las desventajas de las diferentes herramientas que se presentaron en los párrafos anteriores:

Herramienta	Etapa	Desventaja en su uso
Google	Obtención	Se requiere conocer el uso de la opción de búsqueda avanzada.
Youtube	Obtención	Algunos videos están bloqueados o restringidos, lo que delimita la búsqueda.
Storyful MultiSearch	Obtención	Algunas de las redes sociales pueden estar bloqueadas.
Google Trends	Obtención	La búsqueda solo se refiere a búsquedas en Google con cierto límite de tiempo.
HTTrack	Scraping	Se requieren conocimientos previos de programación.
Import.io	Scraping	El sitio se encuentra en inglés, no se puede seleccionar otro idioma.
PHPJava y Python	Scraping	Deben conocerse ciertos códigos de programación.
Zamzar	Cambio de formato	Solo permite convertir archivos de hasta 100 MB.
OCR	Cambio de formato	
CometDocs	Cambio de formato	
Tabula	Cambio de formato	El sitio se encuentra en idioma inglés.
Open Refine	Limpieza	Se requieren conocimientos previos de programación.
Excel	Limpieza	No es un software libre.
R	Exploración	Se requieren conocimientos previos de programación.
Tableau	Exploración	Una de sus versiones tiene costo.
Processing	Visualización	Se requieren conocimientos previos de programación.
D3.js	Visualización	Se requieren conocimientos previos de HTML y CSS.
CartoDB	Visualización	Aunque tiene una versión de prueba, tiene costo.

Tabla 2. Herramientas auxiliares. (Fuente: Elaboración propia)

4.2 Plataformas Digitales para el Aprendizaje

La simbiosis entre tecnología y educación tiene su origen en la Reforma Protestante de la Europa del siglo XVI, con la utilización de la imprenta junto a la teología para fomentar el estudio de la biblia. Un segundo episodio de esta aventura es el de la ilustración, donde se creó el sistema público de educación y la expansión de las instituciones de educación superior. Desde el siglo XIX las virtudes de la educación universal y el progreso en la técnica permitió que a partir de cualquier avance en las tecnologías de la comunicación se pusieran a trabajar inmediatamente en las necesidades educativas de las masas (Haber, 2014).

El uso de la tecnología en línea, proporciona facilidades para el proceso de enseñanza-aprendizaje como son: la inmediatez de las interacciones, acceso a los cursos desde cualquier lugar y tiempo, la capacidad de retroalimentación vía comentarios y la discusión como proceso de construcción de aprendizajes (Dorrego, 2016).

De acuerdo con Peña Borrero en Garduño Vera (2005) la educación a distancia es una metodología caracterizada por la interposición de formas mediadoras entre docentes y estudiantes. Con la finalidad de trascender las limitaciones de tiempo y espacio características de la educación presencial. Los componentes esenciales de este tipo de educación son: una

población masiva de estudiantes, la separación física entre el maestro y el alumno, la necesidad inminente de desarrollar materiales didácticos, el aprendizaje centrado en el usuario y el uso de redes sociales académicas de aprendizaje.

Un aspecto fundamental en la educación a distancia es la creación de escenarios virtuales colaborativos de enseñanza-aprendizaje. Esta tendencia se basa en la idea de comunidades de aprendizaje capaces de crear conocimiento colectivo. El reto en este tipo de comunidades es comprometer a los participantes a involucrarse de forma activa y planificada en un marco teórico y pedagógico compartido (Garduño, 2005).

La enseñanza en línea es un tipo de educación a distancia, en la cual se incluyen cursos a través de correo electrónico, videos o conexión por teléfono o televisión. En cualquier caso, debe cumplirse la condición de no estar en el mismo sitio, tanto los alumnos como los maestros. La enseñanza en línea, se basa en la comunicación mediada por computadora (EMC). Es una fuerza democratizadora porque el participante solo necesita tener acceso a una computadora y conexión a la red (Dorrego, 2016).

La educación en línea implica, importantes cambios en la forma como se diseñan, desarrollan y evalúan estos procesos. Entre sus características se encuentran: la colaboración, la conectividad, que está centrado en el estudiante, no existen restricciones de tiempo y espacio, hay facilidad para crear comunidades, existe

un énfasis en la exploración, además de tener diversas experiencias sensoriales y de autenticidad (Kearsley, 2000).

Una manifestación de la educación a distancia son los cursos masivos en línea (MOOCs). Mismos que son una tendencia en crecimiento en el área de la educación superior. Son experiencias centradas en el usuario, a través de plataformas, instituciones, temas y profesores. Los MOOCs son una nueva tecnología educativa, la cual tiene el potencial de transformar varios aspectos del aprendizaje.

Desde un punto de vista histórico, los cursos masivos en línea son la evolución de experiencias anteriores como: la educación abierta y el *e-learning*. En un curso masivo en línea participan cuatro tipos de alumnos: los auditores o participantes pasivos, los cumplidores o participantes activos, los desconectados e inconstantes y los catadores o moderadores. (Valverde Berrocoso, 2014).

Para los participantes de un curso masivo en línea, la experiencia representa un conjunto de actividades características de la educación tradicional como: lecturas, tareas, reportes. Pero, la diferencia radica en la posibilidad del MOOC de ser atendido por cientos de participantes al mismo tiempo (Haber, 2014).

El desarrollo de un curso masivo en línea, implica una serie de retos pedagógicos. El primero, es balancear una investigación profunda y la creación de conocimiento sofisticado, articular la amplitud y la profundidad de la participación de los estudiantes, diseñar procesos y prácticas

para motivar roles participativos, contemplar estrategias para optimizar la contribución de los docentes y los participantes más avanzados (Valverde Berrocoso, 2014).

Es importante diferenciar entre un curso en línea y un curso masivo (MOOC). El primero, tiene como objetivo desarrollar una plataforma *e-learning* con funcionalidades, así como estructuras acotadas y diseñadas para la interacción directa con un profesor. En el segundo caso, se busca un diseño tecnológico para facilitar la disseminación de actividades hacia los participantes a través de una o varias plataformas. En la tabla 3 se muestra la diferencia entre cada una de las tendencias de enseñanza/aprendizaje arriba mencionadas:

El carácter abierto de un curso masivo en línea MOOC se debe a la disponibilidad de las actividades, los contenidos, los recursos y la accesibilidad para cualquier usuario. Por otra parte, en un curso en línea, la plataforma de aprendizaje no es la clave, sino el debate en torno a los contenidos y la generación del conocimiento por parte de los participantes (Universidad Carlos III de Madrid, 2017).

Una de las fortalezas más importantes de un curso masivo en línea radica en la experiencia positiva de intercambio y colaboración entre los participantes, durante el proceso de enseñanza aprendizaje. Esto se debe al potencial de los elementos afectivos, sociales y cognitivos, los cuales son características del aprendizaje colaborativo en entornos virtuales (Valverde Berrocoso, 2014).

No	Curso en línea	Curso masivo en línea
1	Entorno cerrado	Entorno abierto
2	Acceso Previo. Algunas veces con un pago.	Acceso Gratuito. En algunos casos se debe pagar por una certificación.
3	Cupo limitado	Participación masiva
4	Apoyo directo del profesor	Apoyo de la comunidad
5	Comunicación a través de foros de debate	Diversas herramientas de comunicación, como redes sociales
6	Orientado hacia la evaluación y la acreditación	Énfasis en el proceso de aprendizaje más que en la evaluación y acreditación

Tabla 3. Diferencias entre cursos masivos y cursos en línea. (Fuente: Elaboración propia)

El aprendizaje colaborativo consiste en actividades dinámicas, pues el docente se convierte en un facilitador, la enseñanza y el aprendizaje se convierten en experiencias compartidas, los alumnos son totalmente responsables sobre su proceso de aprendizaje, la reflexión sobre su proceso cognitivo y la construcción de consensos por medio del trabajo en equipo (Kirschner, 2001).

Desde el punto de vista educativo los cursos masivos en línea tienen las siguientes fortalezas: el valor de la información en línea para la educación superior, la creciente demanda de formación universitaria especializada y de calidad, la adaptación de las metodologías docentes a las

características de la sociedad red, la definición de nuevos roles docentes en la educación superior, el uso de estrategias de evaluación para dar al alumno un papel activo, la flexibilidad y adaptabilidad de la oferta educativa, la colaboración entre instituciones de educación superior, el acceso libre a una amplia oferta educativa, el desarrollo de herramientas tecnológicas para apoyar el proceso de enseñanza aprendizaje y la experiencia de colaboración en línea entre estudiantes (Valverde Berrocoso, 2014).

Entre sus desventajas se encuentran: el uso de metodologías de enseñanza basadas en teorías educativas superadas, en la presentación de

algunas supuestas innovaciones educativas, la estandarización del conocimiento, la ausencia de un diseño universal para el aprendizaje, la falta de conocimiento sobre el desarrollo y evolución del aprendizaje en línea, la desvalorización del papel del docente en la evaluación y la tutorización, el rol docente como figura unidireccional, una visión parcial del concepto de educación abierta, preponderancia en la perspectiva económica frente a la experimentación y la investigación y la necesidad de cierto nivel de competencia digital y autonomía en el aprendizaje (Valverde Berrocoso, 2014).

La adopción de la iniciativa MOOC en la región de América Latina ha tenido un incremento notable en un espacio muy breve de tiempo. A septiembre de 2016 se registraron 559 cursos producidos en la región, con un incremento notable de 141 cursos del 2015 al 2016. La intensidad en la producción de cursos en América Latina supera la intensidad que tuvieron las iniciativas MOOCs, en sus inicios en Europa (Pérez Álvarez, Maldonado, Rendich, Pérez Sanagustín, & Sapunar, 2017).

En la región de América Latina la producción de cursos se concentra más en algunos países que en otros. Tres países lideran la producción de MOOCs en la región: México, Colombia y Brasil. La primera posición es ocupada por México con una producción de 157 cursos. Sin embargo, es importante anotar que el 62% de los países han adoptado la producción de iniciativas MOOCs, un porcentaje similar al 64% registrado

en Europa (Pérez Álvarez, Maldonado, Rendich, Pérez Sanagustín, & Sapunar, 2017).

El profesional y/o ciencias aplicadas representan el principal dominio de los cursos producidos en América Latina y el idioma que predomina es el español. Así mismo, para el 2016 se nota un incremento en la producción en las áreas de las ciencias formales y las humanidades. Por otro lado, en Europa predomina el mismo dominio que en América Latina, pero se ha hecho un esfuerzo mayor en la producción de MOOCs en áreas de las Ciencias Sociales. En Europa la mayoría de cursos se producen en el idioma inglés, seguido por el idioma español (Pérez Álvarez, Maldonado, Rendich, Pérez Sanagustín, & Sapunar, 2017)

En la red, existen infinidad de cursos en línea para el aprendizaje de temáticas relacionadas con el manejo de datos. En los párrafos siguientes se analizan las características de seis iniciativas: General Assembly, Data Tells a Story, Big Data University, DataCamp, DataBasic y Tutellus.

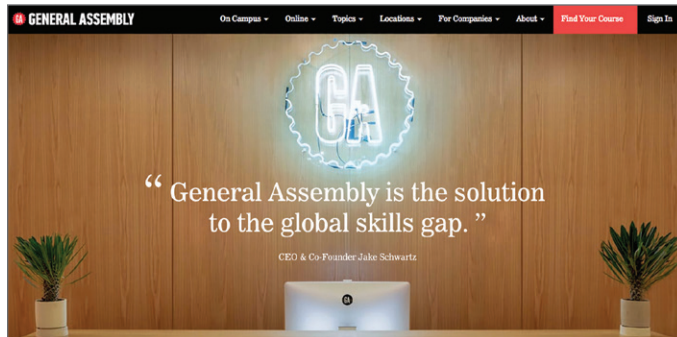


Figura 22. Home de la página principal del sitio oficial de General Assembly.

4.2.1 General Assembly³⁷

Es una compañía encargada de solucionar la brecha de habilidades en diseño, tecnología, marketing y datos a nivel global. Su objetivo es formar una comunidad de profesionales tanto individuales como colectivos, a través de la educación y la conexión estratégica de diversas trayectorias profesionales. Es pionera en educación, la transformación y la especialización en las habilidades técnicas más demandadas de la actualidad. Entre las oportunidades para los individuos al tomar cursos se encuentran: la construcción de habilidades a nivel profesional, la colaboración y la conexión con una amplia red de profesionales y un cambio e impulso a su carrera o portafolio. Con la finalidad de innovar en el entrenamiento y el aprendizaje a lo largo de la vida en las áreas de conocimiento tecnológico. General Assembly proporciona los siguientes cursos: programación, diseño y experiencia de usuario (UX), manejo de datos, mercadotecnia, negocios e impulso

al desarrollo de las trayectorias profesionales. En el área específica de manejo de datos, esta plataforma, proporciona: cursos de tiempo completo, de medio tiempo, clases y talleres, así como eventos.

4.2.2 Data Tells a Story: Reading Data in the Social Sciences and Humanities³⁸

Es un curso ofrecido por la Universidad de Loughborough en el Reino Unido. Su finalidad consiste en mostrar el papel de los datos en diversas disciplinas y las principales herramientas para extraer conocimiento de los datos. Sus principales características son su carácter gratuito, sus dos semanas de duración y la posibilidad de obtener un certificado. Respecto a su contenido se hace una introducción a las ideas básicas sobre la obtención de datos, y las herramientas para transformar datos en información útil.

El curso está dirigido para el público en general con un buen dominio de las matemáticas y habilidades de pensamiento crítico. Pero, principalmente es atendido por profesionales de las artes, las humanidades, las ciencias sociales y el deporte. En general cualquier persona con la suficiente curiosidad para aprender y experimentar como los datos son recuperados y analizados para responder ciertas cuestiones.

El contenido del curso está estructurado en

³⁷ <https://generalassembly.ly>
(Consultado el: 3 de octubre de 2017)

³⁸ <https://www.futurelearn.com/courses/data-tells-a-story>
(Consultado el: 9 de noviembre de 2017)

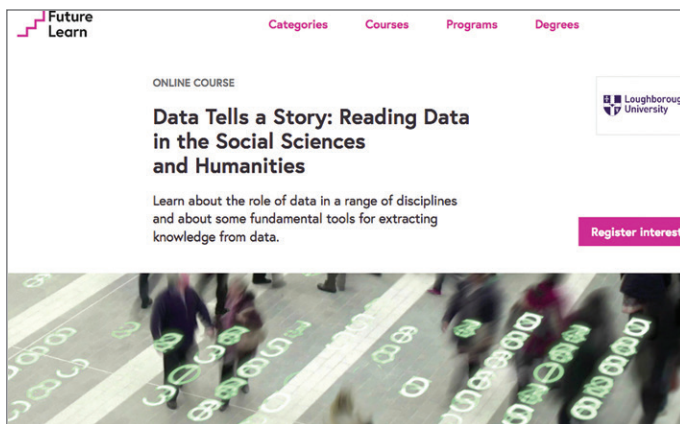


Figura 23. Página inicial del sitio Data Tells a Story.

dos secciones. La primera, muestra algunas preguntas acerca de las artes, la política, la geografía y los deportes. Y las distintas posibilidades para responderlas a partir de datos y la manera de ubicarlos y recuperarlos. A partir de ellos, se utilizan algunas herramientas visuales y se discute como interpretarlas con base en lo que los datos tratan de decir.

La segunda parte del curso aborda las formas de presentar y resumir datos. Es decir, se muestra el camino apropiado para elegir los instrumentos adecuados para mostrar los datos en función de la pregunta inicial y los datos disponibles. El curso finaliza con la revisión de un artículo donde se muestra las formas de comparar datos y explicar críticamente su significado. A lo largo del curso se desarrollan habilidades para recolectar, compartir, analizar y discutir los datos sobre el aprendizaje en el curso y las cuestiones que estos tienen que decirnos.

4.2.3 Big Data University³⁹

La misión de este proyecto es fomentar las ventajas del análisis, la ciencia de datos y los datos masivos para formar recursos humanos capacitados para explotar adecuadamente la ingente cantidad de datos a nivel mundial. Para ello, es fundamental aprender los conceptos fundamentales para desarrollar, procesar y analizar los datos masivos.

Su propuesta educativa está compuesta por un programa en línea a través de un sitio web con acceso gratuito a contenidos y herramientas. Cada curso está diseñado para ayudar al participante a seguir su propio paso con base en sus prioridades. Son dos las razones principales para tomar estos cursos: la familiaridad y experiencia práctica con tecnología innovadora de forma gratuita, y la confianza para construir y contribuir a esta comunidad con habilidades fundamentales para hacer diferencia. Son cuatro

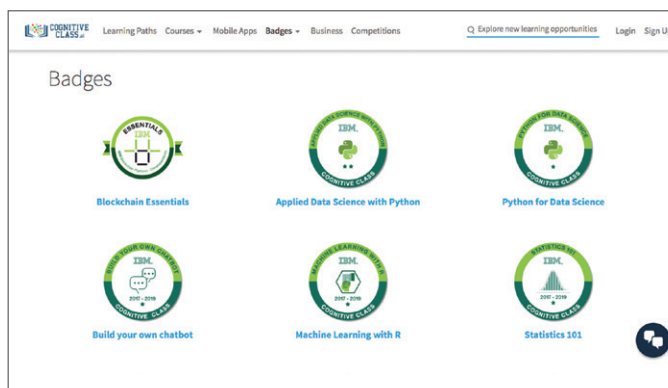


Figura 24. Una de las secciones del sitio de Big Data University.

³⁹ <https://bigdatauniversity.com> (Consultado el: 10 de octubre de 2017)

las características innovadoras de esta iniciativa: seleccionar el camino de aprendizaje, completar el curso, ganar puntos y mostrarlo al mundo.

De igual manera, son 7 los caminos de aprendizaje que pueden seguirse en esta plataforma, como se muestra en la Figura 24. Estos grandes temas son: aprendizaje profundo, fundamentos de ciencia de datos, fundamentos de datos masivos, programación en Scala para ciencia de datos, fundamentos de hadoop, fundamentos de spark y ciencia de datos para negocios.

Una característica interesante de esta plataforma es el uso de puntos y bolsas para obtener una especie de medallas cada que se acredita un curso o camino de aprendizaje (ver Figura 24).

4.2.4 DataCamp⁴⁰

Es un curso diseñado para dominar el análisis de datos desde las necesidades y la experiencia de especialistas en R, Python y la visualización de datos. La esencia de esta iniciativa consiste en proporcionar herramientas para el aprendizaje de las habilidades necesarias para implementar proyectos basados en datos.

Son seis los pilares sobre los que descansa este curso. La primera, el énfasis en el aprendizaje de las habilidades de manipulación, visualización y programación de datos. La segunda, es la posibilidad de agregar lenguajes, técnicas y tecnologías a una caja de herramientas. La tercera, es

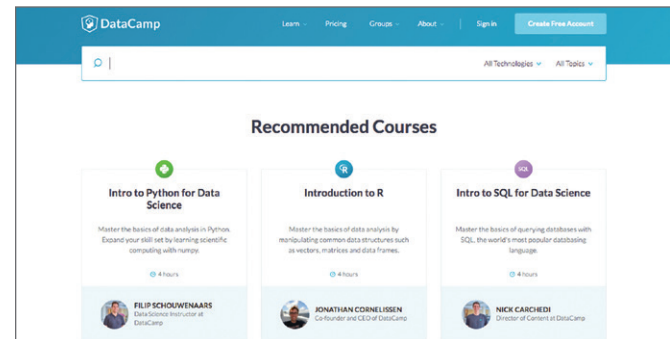


Figura 25. Sección de cursos del sitio DataCamp.

el aprendizaje haciendo, pues no utilizan sólo videos tutoriales sino que se aplican ejercicios de programación específicos directamente sobre el sitio web. La cuarta, es la individualización de los ejercicios, pues permite aprender y codificar al ritmo de cada estudiante. La quinta, es la posibilidad de obtener certificados y comparar los progresos con participantes alrededor del mundo. La sexta, es el ritmo de actualización de la plataforma con los avances más recientes en la ciencia de datos.

4.2.5 DataBasic.IO⁴¹

Es un proyecto del Instituto de Tecnología de Massachusetts dirigido por Rahul Bhargava y Catherine D'Ignazio. Consiste en un conjunto de cuatro herramientas orientadas al tratamiento de datos a nivel principiante. La filosofía es construir herramientas sencillas para facilitar el trabajo con datos de manera divertida.

⁴⁰ <https://www.datacamp.com>
(Consultado el: 19 de noviembre de 2017)

⁴¹ <https://databasic.io/es>
(Consultado el: 14 de octubre de 2017)



Figura 26. Página inicial de DataBasic.

La primera herramienta es *WordCounter*, la cual permite analizar un texto y mostrar las palabras y frases más comunes. La segunda es *WTFcsv* con la cual se puede observar la composición de un documento CSV. La tercera es, *SameDiff* que sirve para comparar dos o más archivos de texto para saber que tanto se parecen. La cuarta es *ConnectTheDots* que muestra cómo están conectados los datos al analizarlos como una red.

Los principios de esta plataforma es su accesibilidad para todos, hacer un proceso de datos bien sin pretender abarcar una amplia gama de tareas, un enfoque basado en el aprendizaje, ajuste a cualquier proyecto a través de la integración de herramientas y también la garantía de la protección de la privacidad de los usuarios y su dirección por educadores.

4.2.6 Tutellus⁴²

Es la plataforma de aprendizaje colaborativo más grande en el sector de habla hispana. Se basa en video cursos de diferentes niveles y temáticas con base en tres principios: el aprender, el enseñar y el promover. En el primero, se ofrece una gran cantidad de temas para aprender, en el segundo es posible demostrar las capacidades docentes y en el tercero es posible compartir cursos con otros usuarios y ganar dinero.

Para utilizar esta plataforma es necesario pagar una suscripción con una tarifa plana, la cual permite tomar ciertos cursos sin preocuparse por el precio individual de cada uno de ellos. Una de las ventajas de esta plataforma es la posibilidad de obtener certificaciones oficiales de la Universidad Rey Juan Carlos y la Universidad Católica de Murcia.

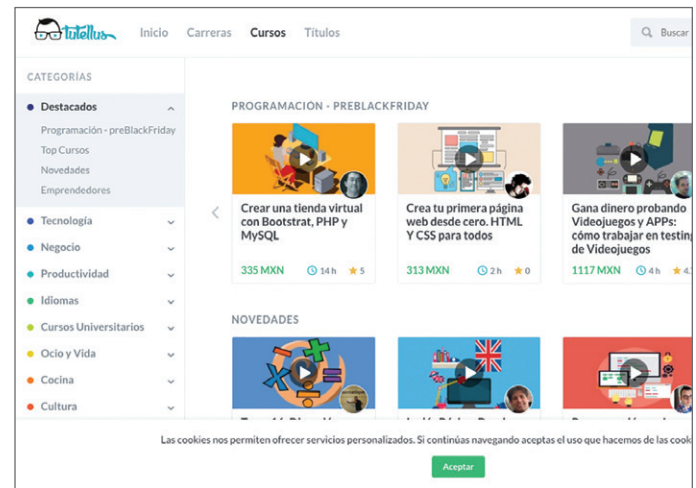


Figura 27. Plataforma de aprendizaje Tutellus.

⁴² <https://www.tutellus.com/>
(Consultado el: 9 de noviembre de 2017)

4.3 Resumen del capítulo 4

Se sabe que cada día se abren más portales para la obtención de D.A. y centros de datos en la red, que sirven como motor de búsqueda para la recopilación de los mismos. Independientemente de que sean gobiernos u organizaciones no lucrativas, ya existe una significativa apertura de datos para el periodista. Ahora solo debe aprender a manipular las grandes cantidades de información que serán la base de su investigación, por lo que su labor será definida por las herramientas que utiliza en cada uno de los pasos del P.D. La elección de las herramientas, sin duda, tendrá que ver con la afinidad que el periodista tenga con ellas pero debe tener el criterio necesario para ello, sin olvidar que muchas de ellas solo están disponibles en idioma inglés y el dominio de este será de gran utilidad.

Se observaron las etapas del proceso y cada una implica ciertos requerimientos específicos para el periodista. Así, por ejemplo, para el proceso de obtención de datos, etapa fundamental para plantear una hipótesis periodística, la búsqueda del periodista debe ser muy específica. O bien, manejar ciertas herramientas o software especializados pensados exclusivamente en ayudar a realizar la búsqueda o que le permitan extraer información, cuando:

- la información contenida en un sitio web no esté disponible para descargar y se vea

en la necesidad de *scrapear* un sitio con el propósito de traducir información contenida en internet a formatos editables, con los que sí le será posible trabajar.

- se encuentre con información en la red que no se puede procesar porque de origen es un archivo de imagen o un archivo *Portable Document Format* (PDF) y

- como en el caso de México, algunas fuentes gubernamentales proporcionen información incluso por medio de fotocopias, lo que hará que el trabajo del periodista se convierta en elaborar bases de datos de manera casi artesanal.

Una vez obtenidos o recuperados los datos, lo normal es que estos presenten espacios en blanco, comillas adicionales, o pueden no estar ordenados. Esta etapa corresponde a la limpieza de datos y aquí el periodista de datos requiere realizar una minuciosa limpieza de los mismos para que sea posible estandarizarlos y establecer una jerarquía dentro de los datos obtenidos.

Posteriormente, una vez que el periodista cuenta con una base de datos ordenada y sistematizada, el siguiente paso es el del análisis de la información, en donde es el periodista con su conocimiento, el encargado de cotejar la información cuantitativa obtenida, con el contexto del fenómeno a indagar.

Ahora se hace necesario que el periodista

entienda la metodología con la que trabajan las herramientas diseñadas para el análisis de datos, apoyándose de un experto en ello. Por supuesto, si el periodista cuenta con conocimientos en matemáticas, informática y estadística, tiene mayores posibilidades de descubrir puntos relevantes dentro de su base de datos.

Por último, solo con ayuda de la visualización, los datos adquirirán mayor sentido para el lector y deben seleccionarse los datos pertinentes de acuerdo a la historia que se desea contar, así como una adecuada representación visual que pueda destacar los datos. En esta etapa también pueden encontrarse otros patrones o conexiones no revelados antes.

Mientras se exploraron algunas de las herramientas existentes para cada etapa del P.D. como parte de nuestra investigación, notamos que desde nuestro proyecto no lograríamos aportar lo suficiente generando una herramienta más para alguna de estas etapas. Mucho más lejana se encontraba la posibilidad de crear una herramienta que se encargara de realizar todas las etapas. Sin embargo, al tratar de entender el funcionamiento de cada herramienta explorada, pudimos notar que en general se asumía que el periodista ya contaba con conocimientos básicos previos para utilizarlas, algo que en la realidad no sucede. Es así como, comenzamos a explorar también las plataformas digitales para el aprendizaje.

Se ha visto que el uso de la tecnología en línea, proporciona grandes facilidades para el proceso de enseñanza-aprendizaje y es capaz

de trascender las limitaciones de la educación presencial, como lo son el tiempo y el espacio y la población masiva de estudiantes.

Desde un punto de vista histórico, los cursos masivos en línea se consideran la evolución de experiencias anteriores como: la educación abierta y el *e-learning*, por lo que analizamos los cursos masivos en línea (MOOCs), que son además una tendencia en crecimiento en el área de la educación superior. Además, en la región de América Latina han tenido un incremento notable en un espacio muy breve de tiempo y las diferencias que presenta con el curso en línea para conocer sus fortalezas y desventajas.

Se presentó también el análisis de las características de seis iniciativas en línea con temáticas relacionadas con el manejo de datos: General Assembly, Data Tells a Story, Big Data University, DataCamp, DataBasic y Tutellus.

El sistema Dataísta es una combinación entre las características de un curso en línea y un curso masivo en línea. De los primeros retoma, el uso de un entorno cerrado de aprendizaje, la comunicación a través de foros de discusión y el registro previo en el sistema para utilizar algunas características. De los segundos, el acceso gratuito al contenido, la finalidad de apoyar a una comunidad y el énfasis en el proceso de aprendizaje más que en la evaluación.

Finalmente, el análisis de las diferentes herramientas y plataformas de aprendizaje nos ha permitido estructurar nuestro sistema y perfilar su objetivo, el cual es desarrollar un Sistema de

Información que permita a los periodistas de investigación entender y utilizar el razonamiento propio de la metodología utilizada en el P.D. para articular sus investigaciones a través de las diferentes bases de D.A.

Una vez que se ha contextualizado el P.D. y se han revisado las herramientas y cursos en línea actualmente disponibles para apoyar a los periodistas en su camino hacia la especialización, es momento de abordar la forma en la que desde una perspectiva interdisciplinaria, se puede contribuir a que el proceso de aprendizaje sea más accesible. Para ello, en el siguiente capítulo se aborda la metodología de investigación aplicada para la realización de este proyecto.

Propuesta Interdisciplinaria

Capítulo **5**

5.1 Interdisciplinarietà

El prefijo *inter* (entre), indica que *entre* disciplinas se va a establecer una relación; determinar el tipo de relación nos conduce a un estudio de los niveles de interdisciplinarietà (Tamayo, 1995).

La Maestría en Diseño, Información y Comunicación (MADIC), de la UAM – Unidad Cuajimalpa, es un posgrado interdisciplinario cuyo objetivo es propiciar el diálogo entre las tres disciplinas que la conforman con la intención de proponer soluciones integrales a problemáticas complejas.

Para entender a profundidad qué es la interdisciplina, es necesario revisar los conceptos de aquellos autores dedicados al estudio de esta propuesta de generación de conocimiento.

Mario Tamayo define la interdisciplinarietà como “un conjunto de disciplinas conexas entre sí y con relaciones definidas, a fin de que sus actividades no se produzcan en forma aislada, dispersa y fraccionada. Es un proceso dinámico que busca proyectarse, con base en la integración de varias disciplinas para la búsqueda de soluciones a problemas de investigación, por lo cual, excluye la verticalidad de las investigaciones como proceso investigativo” (Tamayo, 2004).

Para Morín “el tema de la interdisciplinarietà nació de constatar que la aproximación al mundo a través de una disciplina particular era sesgada y generalmente demasiado limitada, para estudiar una determinada cuestión de la vida cotidiana son precisas múltiples aproximaciones.

A eso se refiere el concepto de interdisciplinarietà” (Hoyos, 2006). La interdisciplinarietà puede significar pura y simplemente que diferentes disciplinas se coloquen en una misma mesa, en una misma asamblea, como las diferentes naciones se reúnen en la ONU sin poder hacer otra cosa que afirmar cada una sus propios derechos nacionales y su propia soberanía con respecto a las intromisiones del vecino... pero la interdisciplinarietà puede significar también intercambio y cooperación, lo cual hace que de ella resulte algo orgánico (Morín, 1998).

Según Nikitina “el diálogo interdisciplinario no solamente se trata de receptividad general hacia puntos de vista alternativos. Implica selección activa y juicio crítico” (Nikitina, 2005).

La interdisciplina es el resultado de un diálogo, un tercer incluido producto de la suma de dos contradictorios reunidos para generar nuevo conocimiento. Existiendo una predisposición a nuevos y diferentes puntos de vista, el diálogo es el primer paso hacia ella.

Entender qué es la interdisciplina, así como mantener la mente abierta nos permitió, primeramente, iniciar el diálogo entre las tres disciplinas que conforman el equipo de trabajo. Posteriormente, nos ayudó a llegar a propuestas de solución a la problemática de interés y a la generación de nuevo conocimiento. Ahora bien, es importante entender de qué forma cada una de las disciplinas que conforman el equipo de trabajo aborda la interdisciplina y de qué forma su experiencia aporta a la propuesta de solución.

5.1.1 Estrategias en Comunicación

La parte central de la línea de Estrategias en Comunicación es la concepción de una forma óptima y adecuada para dar solución a problemáticas que usualmente atienden fenómenos comunicacionales.

El planteamiento de estrategias que permitan el logro de los objetivos de un equipo de trabajo, sin lugar a dudas es la aportación más relevante de la línea de Estrategias en Comunicación al equipo de trabajo. Mediante el diálogo con las otras dos disciplinas que conforman el equipo, la identificación del problema, así como el planteamiento de la solución resultó mucho más sustanciosa, porque todo se derivó del cruce de intereses y experiencias de cada uno, enriqueciendo el proceso de trabajo.



Figura 28. Interdisciplina aplicada al proyecto. (Fuente: Elaboración propia)

5.1.2 Diseño de Información

En el Diseño de Información, se trabaja en la solución de problemas de información (Miiksehaar, 2001). Cuando se trabaja de manera interdisciplinaria se genera un plan de acción para abordar alguna solución que haya codificado cada parte, es decir, cuando dos sujetos tienen la misma información, más no los mismos datos es cuando tienen el mismo modo de orientar su acción. Cuando evocan en común el significado

de su acción y se comunican la misma información (Paoli, 1989). Es cuando el diseño avanza, ya que se habla un solo lenguaje dentro de un entorno colaborativo.

Ya no solo se trabaja con la visión propia del Diseño de Información, sino que se complementa con el trabajo de las otras disciplinas en la propuesta de solución a la problemática abordada, en este caso el *P.D.* Esto, enriquece todo el proceso de trabajo, pues proporciona una solución integral y novedosa, pues no solo se aborda desde una perspectiva disciplinar, sino desde la perspectiva de tres disciplinas diferentes entre sí.

5.1.3 Sistemas para la Interacción

La línea de acentuación de Sistemas para la interacción busca identificar los principales problemas de investigación en las áreas de sistemas interactivos, sistemas de información y redes de comunicación. Así como también analizar los problemas complejos relacionados con la información en organizaciones y proponer soluciones tecnológicas viables. En este caso, los Sistemas para la Interacción permiten poner en marcha Dataísta a partir de la programación de la misma además del intercambio con las otras disciplinas para entender el problema y generar soluciones acordes al usuario final, el periodista de investigación.

5.2 Metodología Aplicada en el Proyecto

La parte metodológica de este proyecto está construida en torno a tres dimensiones, planificadas de forma tal que permitan dar respuesta a las preguntas de investigación planteadas con anterioridad.

5.2.1 Dimensión 1

Se basa en una perspectiva biográfica para responder a la pregunta de investigación *¿quiénes son los principales periodistas de datos en México? ¿Cuál es su postura en materia de Datos Abiertos?* Para darle respuesta se plantea la recolección de evidencia empírica a través del método de la entrevista cualitativa.

La entrevista es el instrumento adecuado para reflejar la perspectiva biográfica de esta parte de la investigación. Pues de esta manera se refleja la experiencia de los distintos agentes involucrados en el *P.D.* mexicano. El análisis de esta información permitió diseñar una tipología de las experiencias de los distintos agentes respecto al *P.D.* y los *D.A.* Al comparar diversas experiencias es posible identificar el componente interdisciplinario del *P.D.*, así como la transición de metodologías de trabajo tradicionales hacia la nueva especialización de los periodistas de datos mexicanos (Creswell, 2014).

El objetivo de las entrevistas es analizar cómo diferentes profesionales se insertan en el P.D. mexicano con un perfil profesional específico y cómo la experiencia interdisciplinaria les ayuda o dificulta para trabajar con otros en la construcción de historias basadas en Datos Abiertos.

Para elegir a los agentes a entrevistar se utilizó como criterio el ubicar a profesionales de diferentes disciplinas involucrados en el P.D. en México como periodistas, diseñadores y programadores. En el transcurso de la búsqueda se identificaron otros perfiles profesionales como físicos, científicos sociales o administradores.

Lista de los profesionales entrevistados en Periodismo de Datos:

No.	Nombre	Medio / Proyecto	Perfil Profesional
1	Israel Piña	¿Quién Compró?	Periodista
2	Tania Montalvo	Animal Político	Periodista
3	Lilia Saúl	Universal Data	Periodista
4	Yosune Chamizo	Animal Político	Diseñadora de Información
5	Sergio Araiza	SocialTIC	Programador
6	Irving Morales	Morlan	Físico nuclear
7	Eduard Martín Borregón	PODER	Periodista

Tabla 4. Personas seleccionadas para entrevistar

5.2.1.1 Hallazgos

Una vez que las entrevistas fueron realizadas y las respuestas fueron analizadas se hicieron algunos descubrimientos:

- Existe un vacío de formación por parte de las universidades donde se imparte la carrera de periodismo, es decir, no se están impartiendo materias relacionadas con el uso de las TICs en el ejercicio de dicha profesión.
- El periodista de investigación, es quien está acostumbrado a trabajar siguiendo una

metodología. Posee la habilidad de identificar los sitios donde puede obtener la información necesaria para llevar a cabo sus investigaciones, así como de contextualizar datos. Es el más afín para integrar la metodología del P.D. a su labor.

- Existen muchas herramientas que son de gran ayuda al momento de implementar las diferentes etapas de la metodología propuesta por el P.D., sin embargo, muchas de ellas asumen que el usuario tiene conocimientos previos en programación, matemáticas o bien, que habla inglés. Situación que complica en muchos casos la incursión del periodista.

■ Se requiere de un lenguaje común con otras disciplinas (Diseño y Programación) para el desarrollo de investigaciones basadas en datos.

■ La falta de tiempo dificulta que los periodistas aprendan la metodología del P.D.

■ La realización de dichas entrevistas nos permitió identificar al Periodista de Investigación como el usuario al cual se enfocará la propuesta de solución a desarrollar. Además, permitió la identificación de la necesidad general a satisfacer; que es la incursión del periodista de investigación al P.D.

5.2.2 Dimensión 2

Se basa en una perspectiva pragmática para responder a la pregunta *¿qué metodología de trabajo utilizan los periodistas de datos mexicanos para construir historias basadas en datos?* Para responderla se recolectó evidencia empírica a través del método de la observación participativa.

El enfoque pragmático se enfoca en el qué y el cómo del problema de investigación y en sus posibles consecuencias. La postura pragmática abre las puertas a múltiples métodos, diferentes cosmovisiones y asumpciones, así como diferentes formas de recolección y análisis de datos (Creswell, 2014).

Tabla 5. Temas abordados durante el “Taller de Datos Extremo” (Fuente: Elaboración propia con base en los contenidos planteados por “Escuela de Datos”)

No	Tema	Ponente	Duración	Participantes
1	Filas y columnas	Sergio Araiza	4 horas	10 Periodistas
2	Extracción / Scraping de Datos	Sergio Araiza	3 horas	
3	Limpieza de Datos	Eduard Martín - Borregón	4 horas	
4	Análisis de Datos y Estadística Básica	Inving Morales (Morlan)	4 horas	
5	Visualización de Datos	Emmanuel Landa (Morlan)	3 horas	
6	User Experience	Sandra Barrón	4 horas	

Por otro lado, la observación participativa es una estrategia de campo, cuyo principal interés es entender el lado humano y las interacciones de un grupo de individuos quienes participan en una configuración específica. Se basa en la localización de situaciones cotidianas del aquí y ahora (Flick, 2009).

En este método la lógica y el proceso de indagación son abiertos, flexibles y oportunos, además requiere de una redefinición constante de las situaciones problemáticas con base en los hechos recolectados en una configuración concreta de la experiencia humana (Flick, 2009).

La observación participativa debe entenderse como un proceso en dos direcciones.

En la primera, el investigador debe ganar acceso al campo y a las personas. En segundo lugar, debe moverse a través de un proceso para volverse más concreta y concentrarse en los aspectos esenciales para la pregunta de investigación (Flick, 2009).

Su aplicación nos permitió observar, analizar y comparar las actitudes, habilidades, problemáticas y áreas de oportunidad en el uso de la metodología del P.D. por parte de un grupo de periodistas de investigación.

La observación se realizó dentro del “Taller de Datos Extremo” impartido por “Escuela de Datos”, mismos que se llevaron a cabo del 2 al 17 de febrero de 2017. A continuación, se muestra los contenidos abordados dentro de dichos talleres.

5.2.2.1 Hallazgos

Mediante la participación activa en los talleres, identificamos los siguientes puntos:

- Poco tiempo para la realización de actividades adicionales (tareas).
- Conocimientos por debajo de lo básico en estadística y en el uso de programas como Excel.
- Poca familiarización con bibliografía relacionada con el P.D.
- No existen equipos interdisciplinarios dentro de sus redacciones.
- No cuentan con un lenguaje común para trabajar con programadores o diseñadores.
- Gran interés por aprender y aplicar la metodología propuesta por el P.D.
- Poco tiempo libre, en general.
- Son más afines a formatos como “video tutoriales” para aprender.
- Concentrados en la aplicación de un paso a paso.

- Alto conocimiento del procedimiento de solicitud de información pública vía transparencia.
- Alta capacidad de contextualizar datos al entorno político-social.

La detección de esta información, nos permitió desarrollar una solución integral adaptada a las características y necesidades del usuario.

5.2.3 Dimensión 3

En esta dimensión se respondió a la pregunta *¿qué características debe tener una herramienta que facilite el aprendizaje de las técnicas de*

obtención, limpieza, análisis y visualización de datos por parte de los periodistas de investigación mexicanos? Esto fue posible, mediante la recolección de evidencia empírica a través del método de grupos de discusión, donde se aplicaron los postulados del Diseño Centrado en el Usuario (DCU) para evaluar la usabilidad del Sistema de Información que se desarrolló basado en los hallazgos de las dimensiones previas.

Los participantes de los grupos de discusión fueron seleccionados al azar utilizando la base de datos de periodistas con la que cuenta Escuela de Datos.

Llevar a cabo grupos de discusión y someter a evaluaciones de usabilidad el Sistema de Información permitió desarrollar ideas sobre el proceso más que sobre el resultado. Pues el debate generado permite obtener ideas importantes sobre el

Tabla 6. Resumen de la metodología del proyecto (Fuente: Elaboración propia)

¿Qué tipo de Sistema de Información puede ayudar a los periodistas de investigación a comprender la metodología utilizada en el Periodismo de datos?	Metodología	Hallazgos
 <p>1 ¿Cuáles son las principales dificultades que enfrentan los periodistas mexicanos de datos?</p>	 <p>13 Entrevistas</p>	<p>Usuario, Barreras, Metodología</p>
 <p>2 ¿Qué metodología de trabajo utilizan los periodistas de datos mexicanos para construir historias basadas en datos?</p>	 <p>6 Talleres = 21 Horas de Observación participativa</p>	<p>Intereses, Preguntas frecuentes, Nivel de conocimientos, Métodos de aprendizaje</p>
 <p>3 ¿Qué características debe tener el sistema de información para facilitar el aprendizaje de la metodología del periodismo de datos?</p>	 <p>Grupos de discusión</p>	

tema de discusión, desde diferentes puntos de vista (Barbour, 2013).

Es necesario hacer mención que, los hallazgos de la tercera dimensión de la investigación, no serán presentados en el presente capítulo; sino en el capítulo 7.

5.3 Usuario

5.3.1 Perfil del Usuario

Con base en la información obtenida en las entrevistas, así como de las diferentes fuentes bibliográficas consultadas, definimos como usuario primario al Periodista de Investigación.

Se definió al periodista de investigación debido a que muchos autores definen al P.D. como una adaptación del Periodismo de Investigación a la época actual, es decir, ahora incluye en sus fuentes de información a los D.A.

A continuación, se presentan las características generales del usuario:

5.3.2 Análisis de Necesidades

El Sistema de Información responde a las necesidades reales del usuario, por lo que su diseño se basó en su contexto y pensando en aquellas tareas


Usuario	Características
 <p data-bbox="968 722 1136 776">Periodista de Investigación</p>	<p>Género: Hombres y Mujeres. Edad: 30 - 45 años. Grado: Licenciatura o Maestría. Lugar de trabajo: Oficina o Casa. Generales: Alto sentido ético, compromiso con la sociedad, poco tiempo disponible, poco conocimiento en lenguajes de programación. Competencias: Familiarizado con la realización de solicitudes vía transparencia, el uso de procesadores de texto, uso de buscadores, detección de patrones y contextualización de sucesos. Áreas de oportunidad: Habilidades matemáticas, manejo de programas como Excel.</p>

Figura 29. Descripción de las características del usuario (Fuente: Elaboración propia)

que por su profesión requiere realizar de manera rápida y efectiva. De acuerdo con la información obtenida en las entrevistas, se puede decir que el periodista cuenta con muy poco tiempo libre, sus investigaciones absorben la mayoría de su tiempo. Además de que en su formación no se integra el uso de las nuevas fuentes de información como los Datos Abiertos para el desarrollo de artículos periodísticos, por lo que deben buscar otras formas de seguir aprendiendo. Lamentablemente los tutoriales, cursos y herramientas disponibles en línea normalmente asumen que quienes los consultan cuentan con conocimientos básicos en lenguajes de programación o estadística, cosa que no siempre ocurre.

Basados en lo anterior, podemos clasificar las necesidades de la siguiente manera:

- **Conocimientos Generales**
 - Lenguajes de programación y formatos de archivos más usados.
 - Estadística básica.
 - Conceptos.
- **Metodología del P.D.**
 - Flujo de trabajo.
 - Ejemplos.
- **Herramientas auxiliares**
 - Categorías de herramientas.

La identificación y la clasificación de las necesidades permitió incluir en el Sistema de Información solamente lo que el usuario necesita, además de diferenciarlo del resto de sitios, cursos y herramientas ya existentes.

5.4 Resumen del capítulo 5

Se han incluido ya algunas y existen múltiples definiciones de la interdisciplina pero todas parecen coincidir en la búsqueda de una interrelación o integración de distintas disciplinas para abordar la solución de un problema en conjunto desde aproximaciones también distintas.

El objetivo de la Maestría en Diseño, Información y Comunicación (MADIC) es propiciar el diálogo entre las tres disciplinas que la conforman con la intención de proponer soluciones integrales a problemáticas complejas.

Estrategias en Comunicación, Diseño de Información y Sistemas para la Interacción, son las tres líneas que conforman el equipo de trabajo. A partir del establecimiento de un diálogo entre las distintas perspectivas y experiencias se han podido originar propuestas de solución y nuevo conocimiento.

Parte del diálogo consistió en plantear tres dimensiones como parte de la metodología aplicada al proyecto y que nos han permitido conocer las respuestas a las preguntas de investigación antes planteadas, además de otros hallazgos.

No fue difícil hallar a los principales involucrados en el P.D. en México, ya que no son numerosos. Sin embargo, en el transcurso de la búsqueda se identificaron otros perfiles profesionales como físicos, científicos sociales o administradores. Todo esto a través de una perspectiva biográfica y del método de la entrevista cualitativa. La primera dimensión nos permitió identificar al Periodista de Investigación como el

usuario al que hemos enfocado la propuesta de solución desarrollada y al mismo tiempo la identificación de la necesidad general: la incursión del periodista de investigación al P.D.

La segunda dimensión, desde la perspectiva pragmática y el método de la observación participativa dentro del “Taller de Datos Extremo”, nos permitió observar las habilidades prácticas para utilizar las etapas de la metodología del P.D. de aquellos periodistas de investigación con conocimientos básicos o nulos en P.D. y la manera en que son capaces de entender y asimilar los conceptos básicos. Y finalmente, la tercera dimensión, mediante la recolección de evidencia empírica a través del método de grupos de discusión, nos permitió evaluar la usabilidad del Sistema de Información cuyo desarrollo se ha basado en los hallazgos de las dimensiones previas.

Como resultado de la metodología aplicada al proyecto y del constante diálogo sostenido por las tres líneas de investigación de la MADIC, logramos identificar y clasificar las necesidades del usuario. Esto permitió incluir en el Sistema de Información, Dataísta, aquello que es necesidad del usuario que pretende incursionar en el P.D., a diferencia de las plataformas de aprendizaje y herramientas existentes exploradas con anterioridad que no toman en cuenta al periodista de investigación y sus características.

En el siguiente capítulo, se describe la propuesta de solución a la que se concluyó después del cruce la información recopilada tanto en la investigación documental como de campo.

Sistema de Información para la Incurción y el Aprendizaje de la Metodología del Periodismo de Datos

Capítulo



6.1 Requerimientos

Una vez que se identificaron las necesidades de los usuarios, llegó el momento de empezar a plantear la forma en la que se abordaría la problemática desde el punto de vista interdisciplinario de la MADIC.

Después de indagar, se llegó a la conclusión de que la forma de abarcar de manera integral la mayoría de las requisiciones de los usuarios, sería el desarrollo de un Sistema de Información. En el sentido más básico, se puede definir a un Sistema de Información como todo aquel conjunto de recursos organizados e interrelacionados dinámicamente que permiten el procesamiento de uno de esos recursos -la información- para que los usuarios tengan acceso a ella (Heredero *et al*, 2006).

Teniendo en mente esa definición, se ideó el diseño y programación de un Sistema de Información al que los periodistas pudiesen acceder para aprender el razonamiento detrás de la metodología propuesta por aquellos que ya hacen Periodismo de Datos, no solo en México, sino en todo el mundo.

La finalidad de que el Sistema de Información contuviera conceptos, es que, mediante su uso, el periodista sea capaz de trascender las herramientas y plantear estrategias para abordar una base de datos con la intención de obtener artículos periodísticos a través de ellas. Además de, proveer a los periodistas de un lenguaje común para trabajar de manera colaborativa con otras

disciplinas que también aprovechan los datos en su labor, como lo son Diseñadores de Información y Programadores.

Mediante el cruce de la información obtenida de las entrevistas, así como de la revisión bibliográfica y los talleres propuestos para los periodistas, se desarrollaron los contenidos que sustentaron al Sistema de Información propuesto, Dataísta.

6.2 Diseño Instruccional aplicado al Sistema de Información

El desarrollo de las nuevas Tecnologías de Información y Comunicación, el auge de las concepciones cognoscitivas del aprendizaje y las necesidades pedagógicas de los últimos años han propiciado un desarrollo acelerado de nuevas formas de interacción en el proceso instruccional que permite la integración de esas nuevas tecnologías al ámbito didáctico con una perspectiva amplia y con mayor eficiencia en lo concerniente al mejoramiento del proceso de aprendizaje (Córdova, 2002).

El Diseño Instruccional es necesario en cualquier modalidad para organizar de una manera sistemática, no solo la enseñanza, sino también el aprendizaje. Los profundos cambios que se han producido a raíz de los avances tecnológicos no dejan de lado la forma como se ha venido diseñando la instrucción; es por ello que los modelos instruccionales de hoy se caracterizan por

ser procesos integrales y holísticos, dialécticos, creativos y flexibles (Romero, 2016).

Puede ser abordado desde diferentes teorías de aprendizaje, pero lo importante es no restringirse ante la posibilidad de integrarlas y tomar los elementos valiosos que cada una de ellas proporciona.

6.2.1 Dataísta desde la Perspectiva Constructivista

Desde la orientación constructivista, *“cada persona construye su propia perspectiva del mundo que le rodea a través de sus propias experiencias y esquemas mentales desarrollados”* (Mergel, 1998).

El objetivo del Dataísta, es preparar al periodista de investigación en la aplicación de sus conocimientos a situaciones reales, nuevas y cambiantes; reforzando el aprendizaje significativo. Este es el que se puede incorporar a las estructuras de conocimiento que tiene el sujeto y que adquiere significado a partir de la relación con conocimientos anteriores.

El aprendizaje significativo, permite que el objeto de estudio adquiera un sentido y un significado particular en cada caso, siempre a partir de una asimilación activa. Es capaz de modificar ideas previas, ampliar la red de conocimientos e incluso, establecer nuevas relaciones entre

conocimientos. Así, consiste en revisar, modificar y enriquecer los esquemas previos y establecer nuevas conexiones y relaciones entre ellos, en definitiva, consiste en construir conocimiento (Sangrà & Duart, 2000).

Cada vez se hacen necesarias nuevas metodologías más abiertas, que ofrezcan herramientas para construir el proceso propio de enseñanza-aprendizaje, opciones complementarias y/o alternativas al aprendizaje formal que fomenten el estudio de manera independiente y autónoma.

Los entornos virtuales de aprendizaje resultan ser ideales, pues los materiales didácticos acostumbran a ser multimedia, es decir, combinan diferentes tecnologías al alcance (textos, registros y videos), desde una perspectiva de máximo provecho pedagógico (Sangrà & Duart, 2000). Además, no es necesario coincidir en el espacio ni en el tiempo para poder desarrollar un proceso de aprendizaje personal que permita a quien lo siga, alcanzar los objetivos que se haya planteado. Las ventajas de este sistema son obvias: la superación de problemas de desplazamiento suscitados por la distancia o por alguna limitación física, cumplir el deseo de muchos adultos de iniciar o reemprender estudios especializados y la posibilidad de continuarlos pese a los horarios laborales.

Las diferentes estrategias de enseñanza-aprendizaje deben conducir a un aprendizaje significativo, a una síntesis personal y propia de los contenidos a partir de una diversidad bibliográfica recomendada, actividades propuestas, bases de datos, debates, foros, etc. (Sangrà & Duart, 2000).

6.3 Prototipo

Basado en la revisión bibliográfica, así como en la información proporcionada por los usuarios, se identificaron la mayoría de los problemas que los periodistas de investigación enfrentan cuando tratan de aplicar la metodología propuesta por el P.D. en sus investigaciones. En la hipótesis, se plantea probar que la falta de sustento teórico, así como la falta de conocimientos sobre procesamiento de datos por parte de los periodistas de investigación, son los principales obstáculos por los cuales los periodistas no utilizan los Datos Abiertos en sus investigaciones. Por ello, la creación de un Sistema de Información basado en sus necesidades, les permitirá aprender la metodología del P.D., para así, ser capaces de contar historias a través de los datos.

El prototipo contiene conceptos básicos sobre cada paso de la metodología, mediante el uso de videos infográficos. Así, el usuario es capaz de identificar la forma en la que todas las etapas de la metodología están conectadas. Algunos temas, contienen infografías con el paso a paso sobre cómo realizar cierta actividad, por ejemplo, el *scraping*. Además, cada parte de la metodología tendrá su propia herramienta auxiliar, misma que servirá al usuario para practicar los conocimientos y habilidades adquiridos dentro del sistema.

Uno de los hallazgos dentro de la investigación, fue el saber que los usuarios querían conocer más herramientas en línea que pudieran simplificar sus labores. Por ello, dentro de cada

etapa de la metodología se incluyó un listado de aquellas herramientas más usadas por los Periodistas de Datos.

Además, se incluyó un foro donde el usuario podrá escribir sus dudas e interactuar con otros usuarios del sistema. Se llegó a esta decisión porque durante la observación participativa, se identificó que esta comunidad es pequeña y comprometida con ayudar a otros durante su camino a la especialización.

6.3.1 Mapa del Sistema

Para identificar las secciones que tiene el Sistema de Información, Dataísta, se planteó un mapa del sistema, mismo que permitirá tener claro que secciones lo conforman y los contenidos que se desarrollaron para sustentarlo.

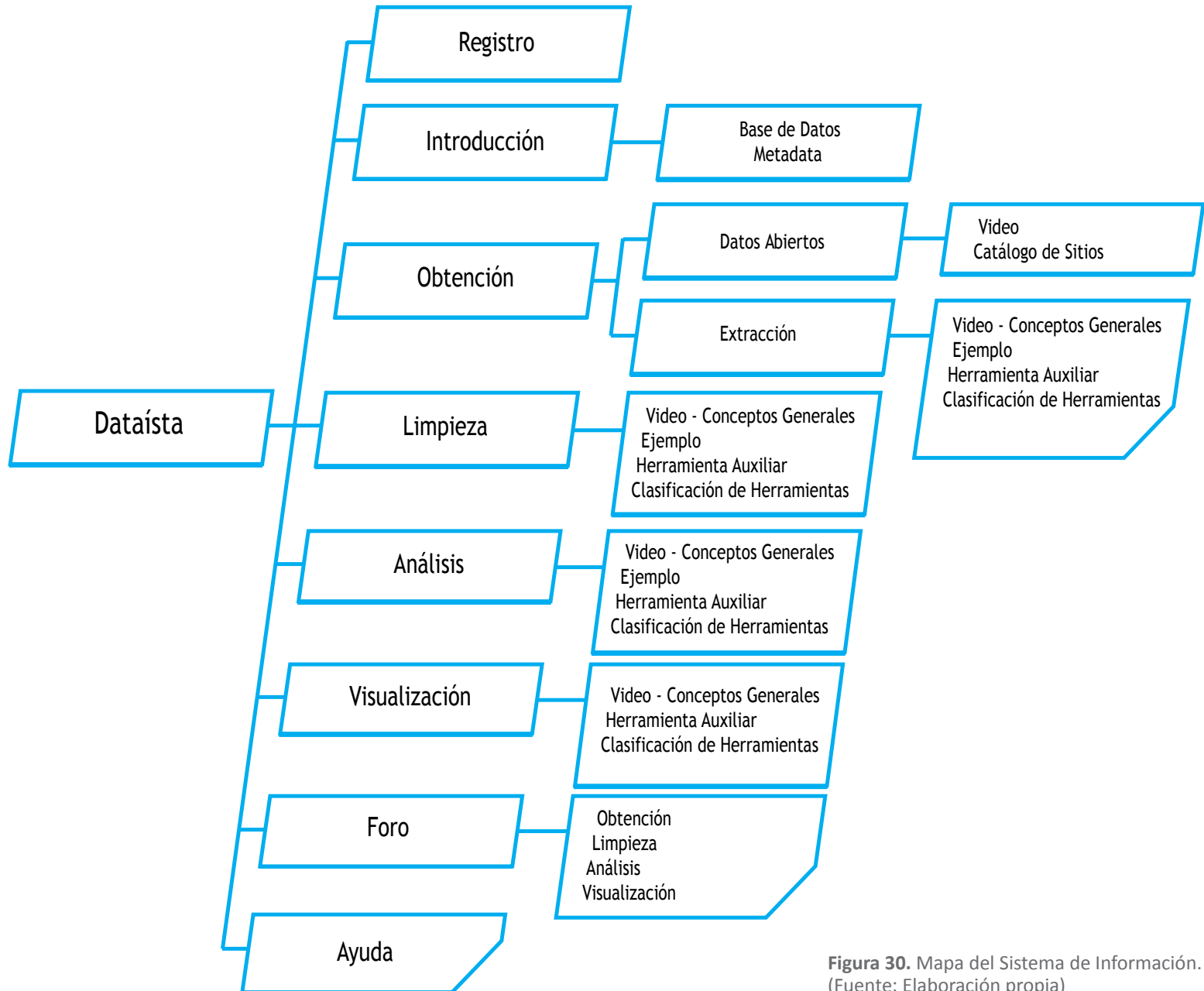


Figura 30. Mapa del Sistema de Información. (Fuente: Elaboración propia)

6.3.2 Modelado del Prototipo

Una vez que se identificaron las características que debía poseer el prototipo de Sistema de Información, el siguiente paso fue diseñarlo. Para el diseño, se decidió implementar los postulados del “Diseño UML”.

El Lenguaje Unificado de Modelado (UML) es una técnica para la especificación de sistemas en todas sus fases (Eriksson, 2004). Su función es la de facilitar a los desarrolladores la identificación de requerimientos del sistema.

El lenguaje de modelado es la notación (principalmente gráfica) que usan los métodos para expresar un diseño. El proceso indica los pasos que se deben de seguir para llegar a un diseño (Análisis y Diseño Orientado a Objetos, 2017).

Generalmente, el Diseño UML contempla la realización de 4 diagramas. Para el prototipo del sistema, se decidió implementar solamente el Diagrama de Casos de Uso y el Diagrama de Actividad.

6.3.2.1 Diagrama de Casos de Uso

Como primer punto del análisis del sistema, se realizaron los siguientes diagramas de uso. En ellos, se pueden observar el tipo de interacciones que el usuario puede tener con el sistema.

La elaboración de los diagramas mostrados anteriormente, permitieron definir el tipo de interacciones que el usuario puede tener con el sistema, esto se vincula de manera directa con las secciones establecidas para el sistema previamente. Asimismo, nos permitió definir que era necesario que el sistema cuente con un administrador; si bien el sistema no alberga bases de datos para las investigaciones, es necesario que se gestione la información proporcionada por el usuario durante su registro y en sus interacciones con el sistema.

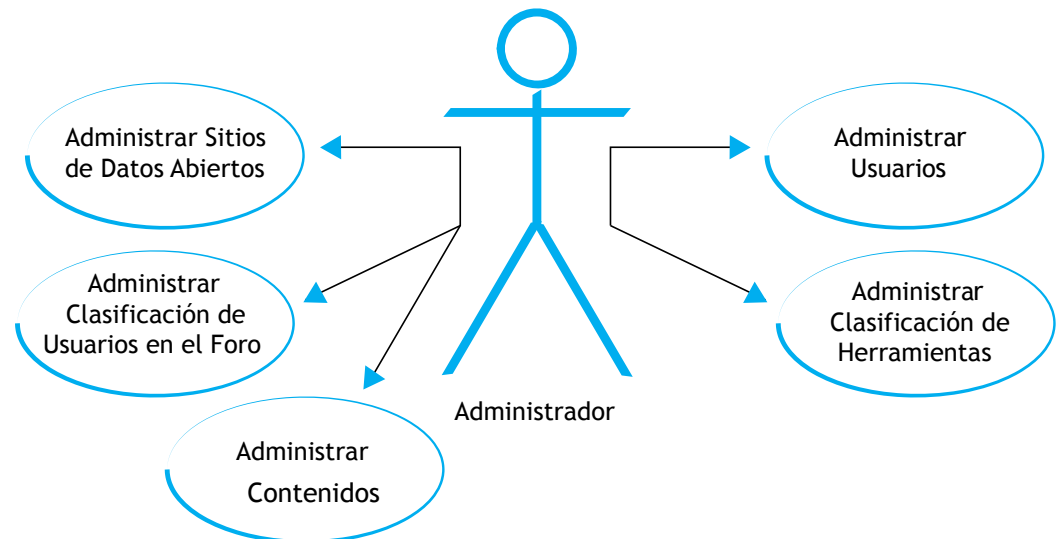


Figura 31. Diagrama de Casos de Uso – Administrador.
(Fuente: Elaboración propia)

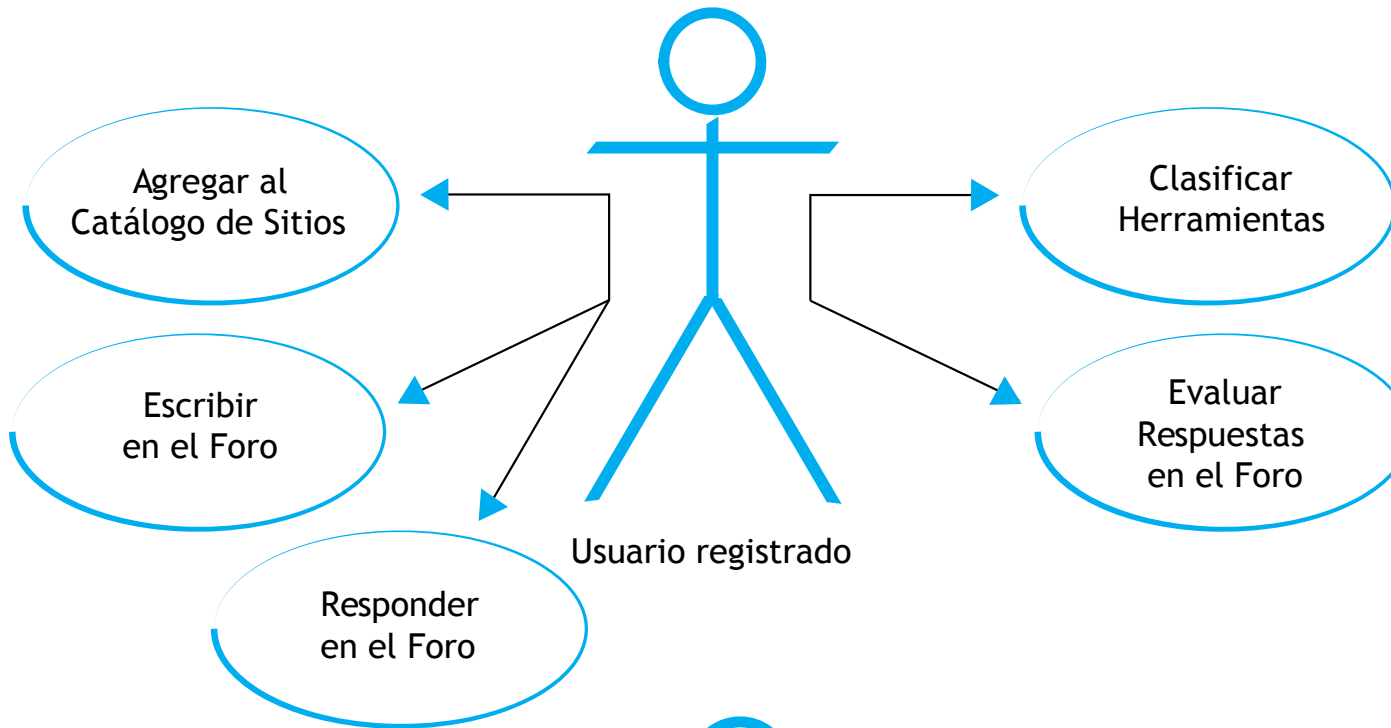


Figura 32. Diagrama de Casos de Uso – Usuario Registrado. (Fuente: Elaboración propia)

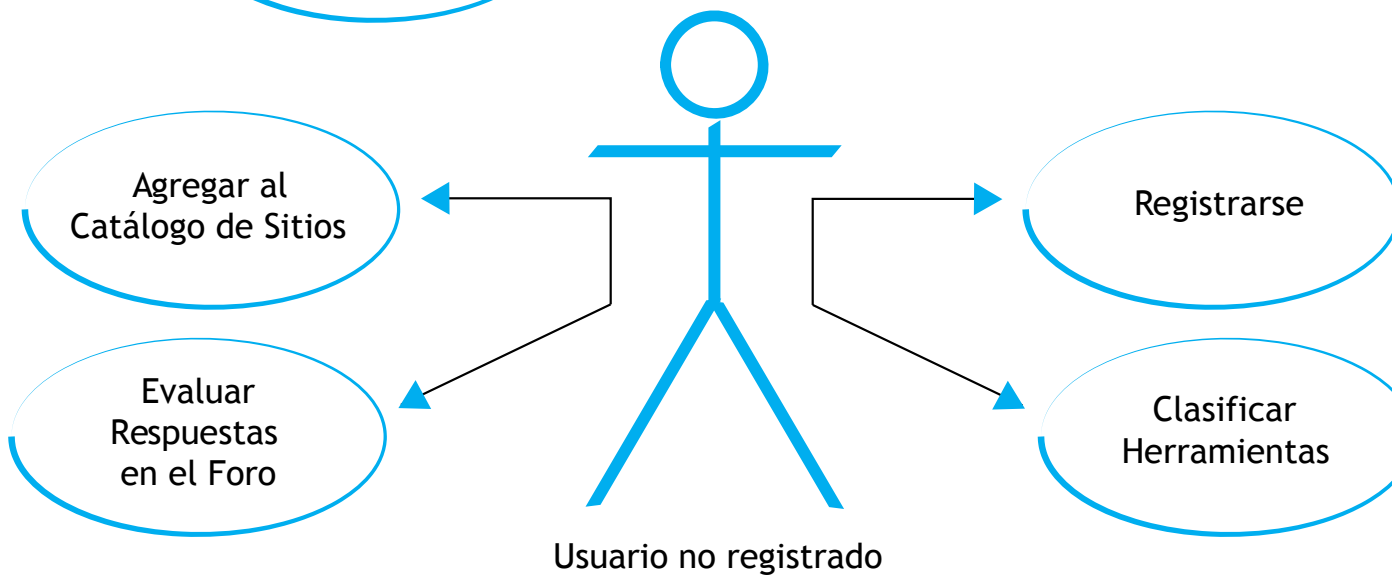


Figura 33. Diagrama de Casos de Uso – Usuario no Registrado. (Fuente: Elaboración propia)

6.3.2.2 Diagrama de Actividades

Este tipo de diagramas, permiten identificar el proceso del sistema, es decir, el flujo de trabajo a través de una serie de acciones. A continuación, se muestran aquellos desarrollados para el prototipo:

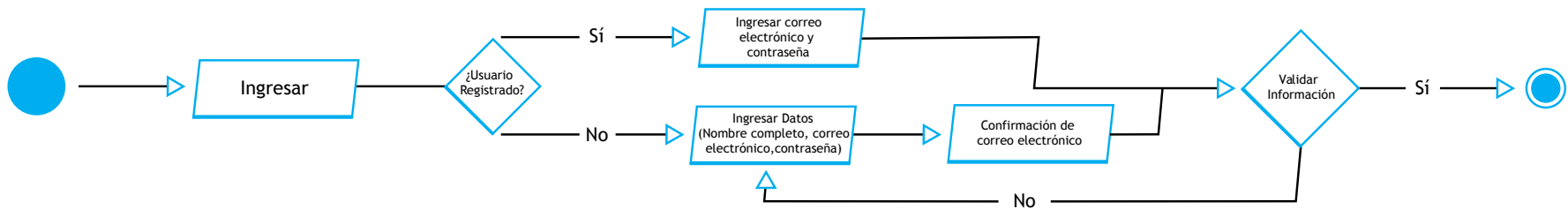


Figura 34. Diagrama de Actividad – Ingresar al Sistema. (Fuente: Elaboración propia)

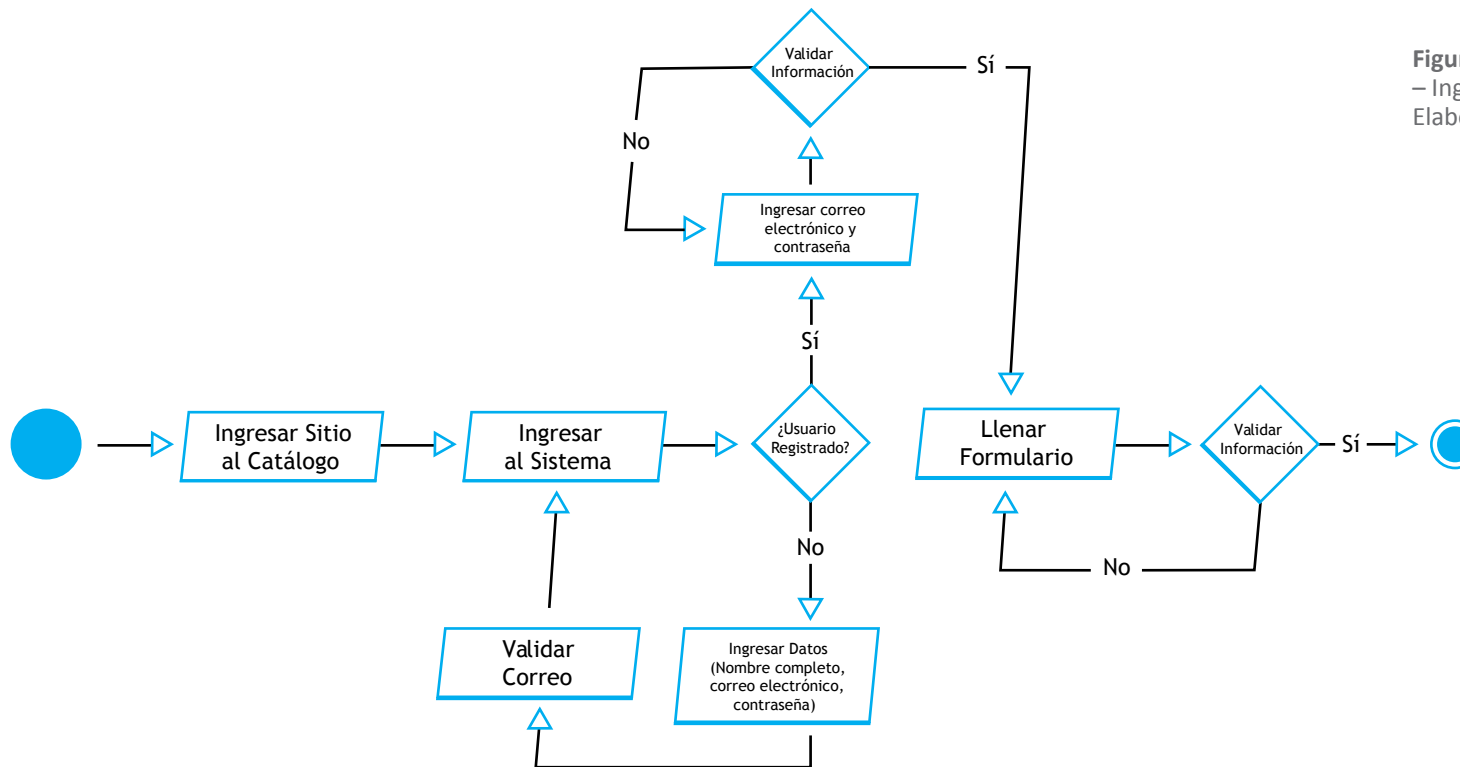


Figura 35. Diagrama de Actividad – Agregar sitio al catálogo de BBDD. (Fuente: Elaboración propia)

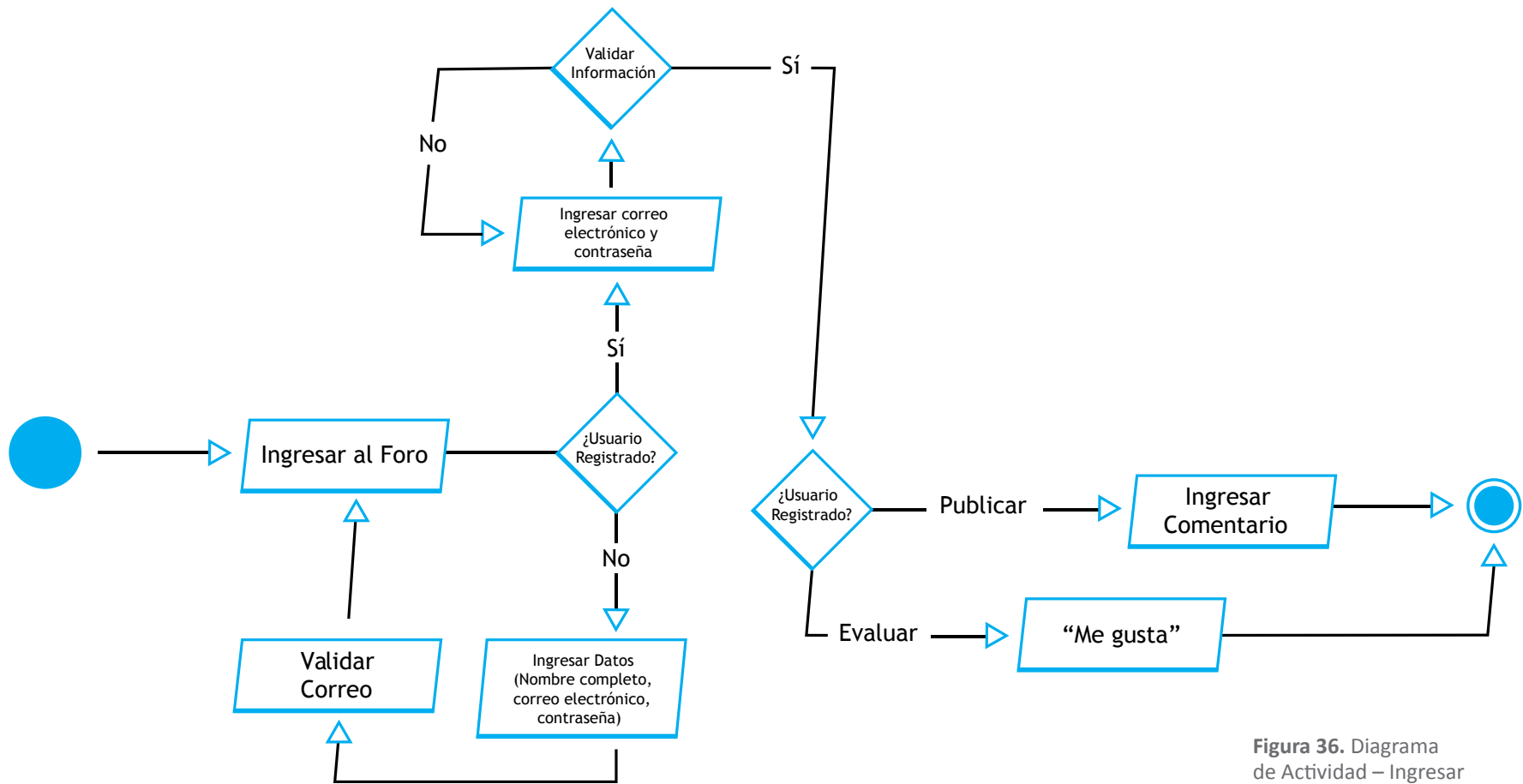


Figura 36. Diagrama de Actividad – Ingresar información al foro.
(Fuente: Elaboración propia)

Los diagramas mostrados anteriormente, permitieron que al momento de empezar a desarrollar el prototipo, se tuviera en cuenta el flujo que el usuario debe seguir al interactuar con el sistema. Esto contribuyó a que cada uno de los pasos fuera considerado y la experiencia del usuario fuera satisfactoria.

6.4 Dataísta

6.4.1 Concepto

Para definir la identidad que tiene el Sistema de Información, fue necesario desarrollar un *brief creativo*, mismo que sirvió como guía durante todo el proceso para tener en mente lo que se pretendía lograr.

A continuación se presentan los puntos clave del *brief* desarrollado.

6.4.1.1 ¿Qué es?

Es un Sistema de Información que surge a partir de las necesidades de los periodistas de investigación por aprender las técnicas relacionadas con el P.D. (obtención, filtrado, análisis y visualización).

6.4.1.2 Propósito del Sistema

El Sistema de Información existe para:

- Albergar información esencial que facilite el entendimiento de cada una de las etapas de la metodología propuesta por el P.D.

- Ser el vínculo entre el conocimiento de aquellos periodistas que ya hacen P.D. y los que aún no, pero están interesados en incursionar en él.

6.4.1.3 Personalidad

Tomando como referencia el objetivo del Sistema, así como las características del usuario, se definieron las siguientes cuatro características como aquellas que definirían la personalidad del Sistema de Información:

- **Profesional**, por las características de la información contenida en el sistema, es importante tener una imagen profesional, para que el usuario no desconfíe de la veracidad de los contenidos.
- **Lógico**, pensando en que se desea transmitir el razonamiento detrás de la metodología, es importante que el usuario tenga la sensación de que existe una lógica que sustenta los contenidos vertidos en el sistema.
- **Amigable**, a pesar de ser un sistema con información conceptual, es importante que el usuario no tenga la percepción de que será un proceso estructurado y rígido.
- **Lúdico**, al ser enfocado principalmente en

el aprendizaje es importante relacionarlo con lo lúdico, pues será un proceso donde el usuario podrá poner a prueba lo aprendido a lo largo de su navegación entre etapas de la metodología.

6.4.2 Conceptualización del nombre

Una vez definidos los principales ejes a comunicar en la identidad del Sistema de Información, el siguiente paso era darle un nombre. De ello surge la idea de utilizar la palabra “data”, relacionada de manera directa con las temáticas abordadas en el Sistema de Información y de agregar el sufijo “-ista”, utilizado para dar nombre a personas que normalmente realizan algún tipo de oficio o profesión. A esta fusión se le conoce como metaplasmo, que consiste en fundir dos palabras en una, es decir, la construcción artificial de nombres mediante iniciales o fragmentos de palabras.

De la mezcla de ambas palabras se llegó a la palabra “Dataísta”, que dio nombre al Sistema de Información. “Dataísta”, hace alusión por una parte a los periodistas, quienes utilizan los datos como parte de su profesión y, por otro lado, a los Datos Abiertos, que están inmersos de manera directa en el contexto del P.D.

6.4.3 Logotipo

Una vez que se tuvieron claros los aspectos a reflejar en la identidad del sistema, así como el nombre que lleva, el siguiente paso fue la elaboración del logotipo. Se decidió que a nivel lingüístico se representaría a qué se dedica el sistema, mientras que la personalidad se vería reflejada a nivel gráfico (Figura 37).

Se definió que el logotipo a desarrollar sería de tipo tipográfico estándar, es decir, el nombre del sistema está escrito de manera normal con alguna familia tipográfica preexistente y de uso libre. La tipografía seleccionada fue “Reality Hyper Regular”. A continuación, se muestran los 3 bocetos seleccionados para la evaluación por parte de los usuarios:

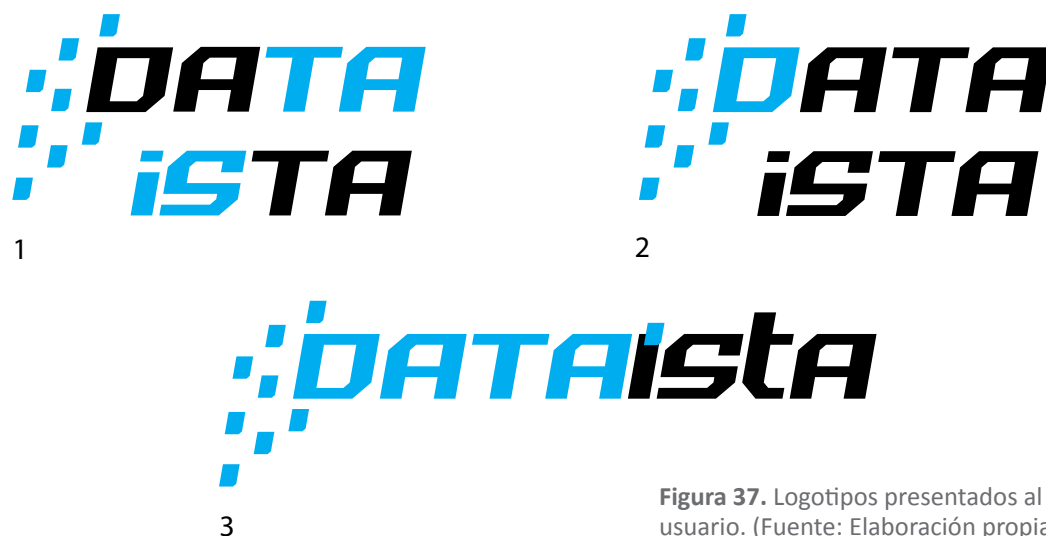


Figura 37. Logotipos presentados al usuario. (Fuente: Elaboración propia)

Además, cuenta con un accesorio estable, esto con la intención de incrementar la capacidad identificadora a través del reforzamiento de algunos aspectos técnicos (pregnancia, diferenciación, llamado de atención). Dicho elemento está representado por algún elemento visual, para

este caso, se decidió que se trata de hacer alusión a los datos.

El logotipo seleccionado fue el número 3 (ver Figura 37), aquí algunas de las razones que motivaron a los usuarios a seleccionarlo:



Figura 38. Nube de palabras de la opinión de los entrevistados. (Fuente: Elaboración propia)



6.5 Diseño y arquitectura del sistema

El diseño y la estructura del sistema de Información, Dataísta, se planteó conforme a un proceso constituido por cada uno de los siguientes puntos, que se siguieron paso a paso para poder desarrollar la propuesta del diseño del sitio.

6.5.1 Objetivos del sitio

Para poder establecer una propuesta de diseño se definieron los objetivos del sitio para que funcionaran como la base o punto de partida en el desarrollo del sitio.

Estos son los objetivos que se consideraron primordiales a tomar en cuenta para definir el sitio:

- Navegación que permita a los usuarios conocer la metodología del P.D. y los conceptos que conforman cada una de las fases.
- Identificar las herramientas disponibles en línea que pueden consultar y utilizar como apoyo en cada una de las fases.
- Uso de la herramienta auxiliar que permita reforzar los conceptos adquiridos mediante la visualización del video.

- Compartir información e interactuar con otros usuarios dentro del sitio.

Con estos objetivos se definieron las características que definen e identifican al sistema.

6.5.2 Bocetaje

Después de tener claro los objetivos del sistema el siguiente paso fue realizar un ejercicio de bocetaje, conforme a los objetivos del mismo. Un boceto es un esquema básico y simple como primer elemento del desarrollo de un proyecto. A continuación se muestra un ejemplo de uno de la serie de los primeros bocetos del sistema.

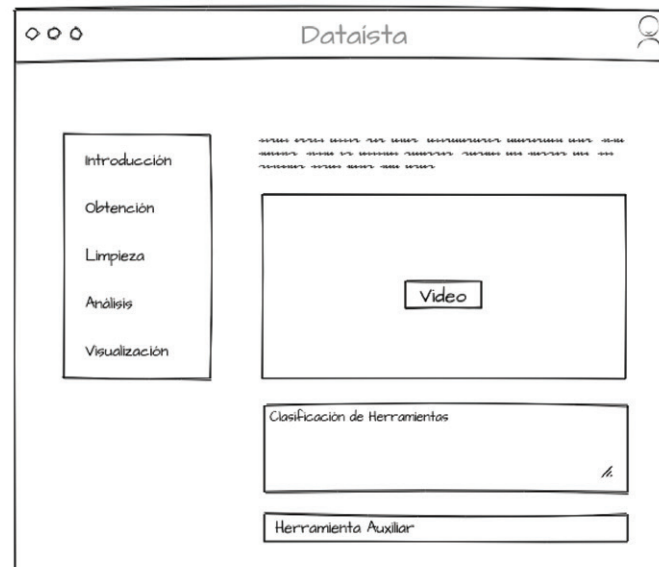


Figura 39. Primer boceto de Dataísta. (Fuente: Elaboración propia)

Después se realizó un primer prototipo con el programa Balsamiq Mockups⁴³ para poder visualizar qué elementos podían constituir al sitio, cómo representarlo y poder entender la navegación, ejemplos:

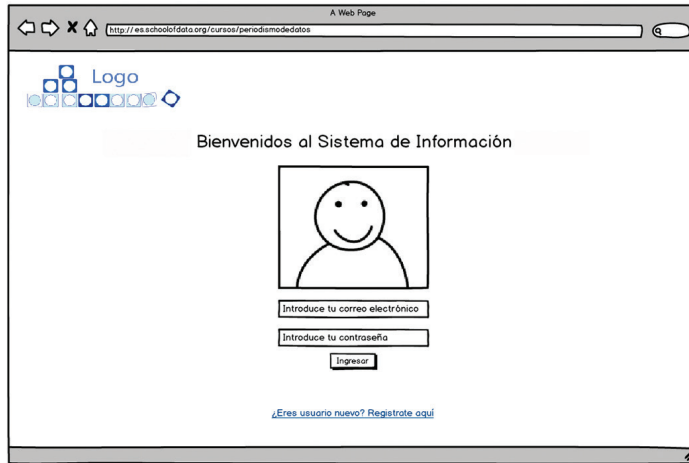


Figura 40. Prototipo de página de inicio de sesión (Fuente: Elaboración propia)

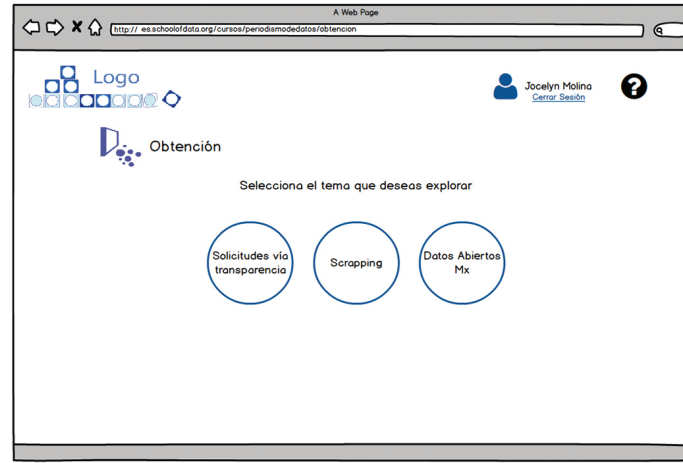
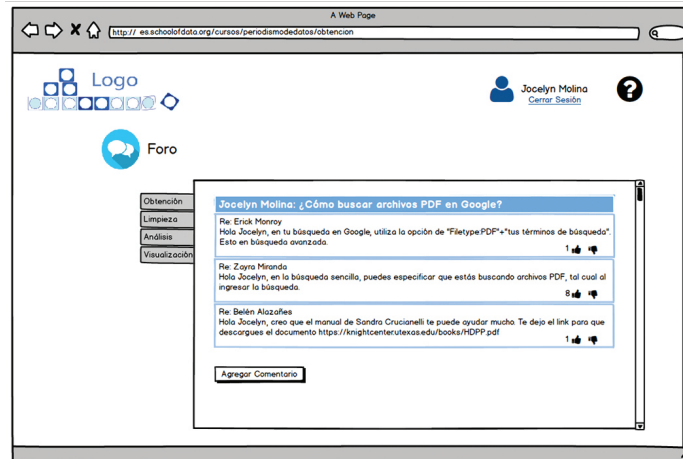


Figura 41. Prototipo del menú de navegación (Fuente: Elaboración propia)

Figura 42. Prototipo de foro. (Fuente: Elaboración propia)



El desarrollo del prototipo permitió definir la funcionalidad del sitio y sobre todo corroborar que cumpliera con los objetivos planteados al inicio. Con esta información, se inició el proceso de diseño de cada uno de los elementos que conforman el sistema.

⁴³ <https://balsamiq.com/products/mockups/> (Consultado el: 3 de octubre de 2017)

6.5.3 Arquitectura del sitio

La Arquitectura de Información como la definiría Richard Saul Wurman (2001, pág. 93) es el estudio de la organización de la información que tiene como objetivo, permitir al usuario encontrar su vía de navegación hacia el conocimiento y la comprensión de la información. Conforme a esta definición diseñamos la estructura de Dataísta tomando como eje central la experiencia del usuario para que este navegue en el sistema entendiendo y comprendiendo toda la información.

Se desarrolló un listado de temas que se consideraron relevantes para la estructura del sitio, misma que se centró en la metodología del P.D.:

Estos temas son los que funcionan como puntos de partida dentro del sitio como un menú de navegación. Esto es similar a un menú interno en donde el usuario puede navegar de una fase a otra sin necesidad de seguir un orden. Establecimos que uno de los criterios más relevantes para

las categorías del sitio y su relación jerárquica eran estas secciones o rangos de la metodología del P.D.

Para dar garantía al sitio de un diseño lógico, coherente, accesible y funcional, en cada fase se decidió incluir un video con los conceptos básicos sin necesidad de poner textos largos y conforme a las necesidades del usuario. En este caso, los videos son una herramienta considerada de fácil uso.

Para complementar estos temas se añadieron otras ventanas que contienen la información que se desarrolla dentro de las fases. A su vez, funcionan como otras categorías que se pudieron agrupar dentro de un menú:

- Medios
- Datos Abiertos
- Preguntas Frecuentes

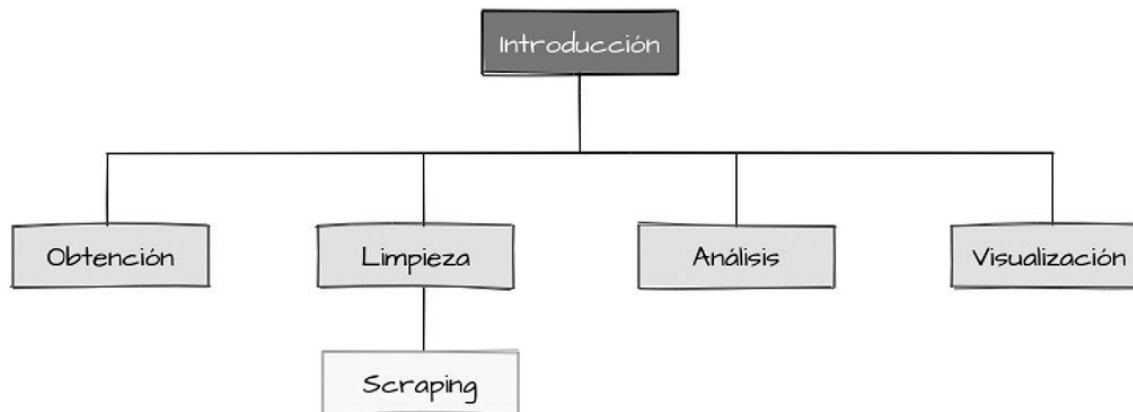


Figura 43. Listado de temas a incorporar en el menú principal del sistema. (Fuente: Elaboración propia)

6.5.4 Diseño de la Interfaz

El diseño del sitio muestra la identidad del proyecto Dataísta: Profesional, Lógico, Lúdico y Amigable. Por lo cual, se diseñó pensando en que debe ser limpio, fácil de recordar, sencillo, sin mucho texto, atractivo y amigable con el usuario.

Se tomó en cuenta el tamaño del monitor de 17 pulgadas con una resolución de 1024x768 pixeles. De igual manera, se consideraron los de un monitor de 640x480 pixeles para poder desarrollar una interfaz acorde al tamaño de los monitores y considerando el acomodo de todos los elementos.

- Se consideró una retícula dentro de las medidas de 5 columnas para el diseño de la interfaz.
- Procurando que fueran pocos los elementos que constituyeran cada parte del sitio para que este pudiera desplegarse en un corto tiempo. Se comenzó por una página principal, o *Home*, en donde un video da la introducción al usuario al sitio.
- Se diseñaron botones simulando un menú en el centro de la página para que sirviera como el elemento de navegación primordial para que el usuario pueda recorrer cada una de las fases de la metodología.
- Se hizo la propuesta de un menú del lado izquierdo para que fuera más sencillo para el usuario navegar y del lado superior



Figura 44 . Retícula del sistema. (Fuente: Elaboración propia)

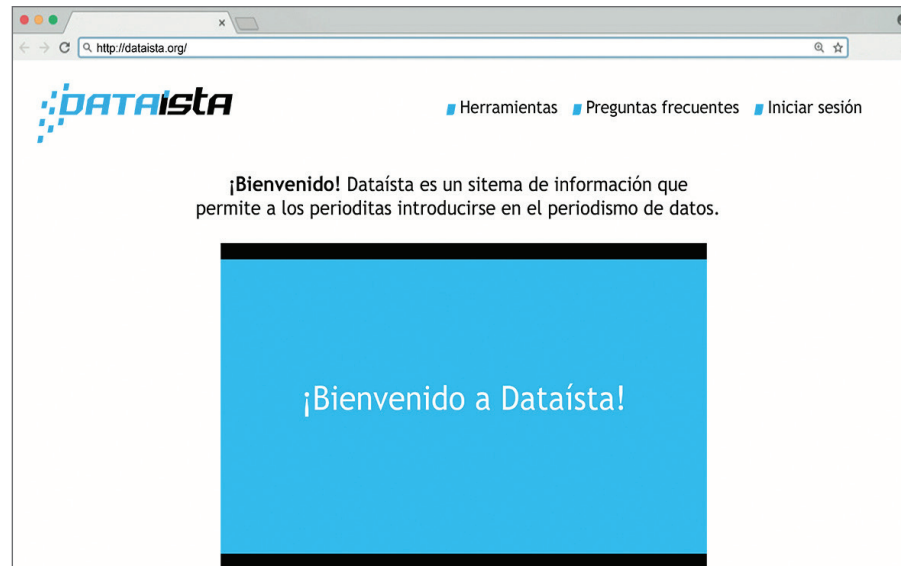


Figura 45. Diseño del home. (Fuente: Elaboración propia)

derecho un menú secundario que ayudara a que el usuario tenga la posibilidad de navegar dentro del sistema en todo momento.

e. Se diseñaron varias propuestas para el acomodo en la clasificación de las herramientas, tomando en cuenta que cada usuario puede calificarlas.

6.5.5 Color

El color del sitio consolida el esquema de navegación, la gama de colores que se eligió para la interfaz corresponde a la identidad que se desarrolló para Dataísta. De esta manera, es el color azul el que predomina dentro de un fondo blanco que detona limpieza y que ayuda a que en el sitio no se presente una saturación de colores o de información.

En algunos componentes se tomó en cuenta algún juego de color o este sirvió para dar jerarquía o rango a algunos elementos.

■ En los textos algunos encabezados tienen color para resaltar a diferencia de los demás elementos.

■ En el menú de lado izquierdo el cambio de sección se interpreta por el color del botón.

■ En la clasificación de herramientas se utilizó un rango distinto de azul para que el nivel de cada grupo fuera más sencillo de identificar.



Figura 46. Botones del menú principal. (Fuente: Elaboración propia)



Figura 47. Menú vertical. (Fuente: Elaboración propia)

6.5.6 Tipografía

En Internet todos los usuarios de la web cuentan con un conjunto básico de fuentes con el fin de asegurar que el contenido de los sitios sea visualizado adecuadamente.

Se seleccionó la tipografía Trebuchet MS, lanzada por Microsoft, que creó la iniciativa Core fonts for the Web en 1996. Esta fuente se encuentra en diferentes sistemas operativos, lo cual va a permitir que los usuarios visualicen sin problema el contenido del sistema.

6.5.7 Botones, menú y elementos de la interfaz

Algunos botones se diseñaron conforme a iconos para representar fases de la metodología. Otros son diseños predeterminados de Wordpress pero que se seleccionaron conforme a los rangos de la identidad de Dataísta para que tanto el diseño de la interfaz como de los elementos que la conforman representaran en todo sentido las características de la identidad.

6.6 Usabilidad

La usabilidad es un atributo que se califica por diferentes factores, la interfaz, navegación, el tipo de usuario y la interacción del usuario con el sistema.

Obtención
Limpieza
Análisis
Visualización

¡Bienvenido a Dataísta!

Obtención. La noción de recolección refiere al proceso y el resultado de recolectar (reunir, recoger o cosechar algo). Un dato, por su parte, una información que permite generar un cierto conocimiento.

Herramientas			
Nombre	Link	Descripción	Evaluación
OPEN REFINE	http://openrefine.org/	App que permite limpiar y transformar datos	☑️ □ □ □ □
EXCEL	https://products.office.com/en-us/excel	Permite limpiar y transformar datos a través de filtros	☑️ ☑️ ☑️ □ □

La noción de recolección refiere al proceso y el resultado de recolectar (reunir, recoger o cosechar algo). Un dato, por su parte, una información que permite...

@Claudia
La noción de recolección refiere al proceso y el resultado de recolectar (reunir, recoger o cosechar algo). Un dato, por su parte, una información que permite...

@Lupe
La noción de recolección refiere al proceso y el resultado de recolectar (reunir, recoger o cosechar algo). Un dato, por su parte, una información que permite...

@Carlos
La noción de recolección refiere al proceso y el resultado de recolectar (reunir, recoger o cosechar algo). Un dato, por su parte, una información que permite...

@Andrés
La noción de recolección refiere al proceso y el resultado de recolectar (reunir, recoger o cosechar algo). Un dato, por su parte, una información que permite...

Figura 48. Acomodo de clasificación de herramientas. (Fuente: Elaboración propia)

Metodología – Dataísta

dataísta.org/metodologia/

Medios Preguntas Frecuentes Iniciar Sesión Home

Periodismo de Datos

Selecciona y da clic a la parte de la metodología que te gustaría explorar.

Introducción Obtención Limpieza Análisis Visualización

Figura 49. Vista del diseño del sistema. (Fuente: Elaboración propia)

La Organización Internacional para la Estandarización define a la usabilidad como:

“La medida en la que un producto se puede usar por determinados usuarios para conseguir objetivos específicos con efectividad, eficiencia y satisfacción en un contexto de uso especificado” (ISO/IEC 9241).

Para poder dar seguimiento a los puntos más importantes a considerar al navegar en el sistema, se desarrolló un cuestionario que evaluó lo siguiente:

- La efectividad del sitio.
- La funcionalidad de la navegación dentro del sitio.
- Que el tiempo para realizar una tarea fuera poco.
- Si el usuario encontraba errores.
- Si los enlaces eran correctos.
- Que los elementos visuales fueran comprensibles.

La usabilidad es un concepto relacionado con la facilidad de uso, es decir se atribuye a la rapidez con que se puede utilizar algo, específicamente se refiere a la eficiencia de utilizar algo⁴⁴.

6.7 Conceptualización y generación del contenido del sistema

Una vez que se definió la identidad gráfica de lo que sería el Sistema de Información “Dataísta”, el siguiente paso fue establecer qué contenidos estarían disponibles en él.

6.7.1 Selección del formato

Como se describió en apartados anteriores se identificó mediante la aplicación de encuestas que el usuario final de “Dataísta”, es decir, los periodistas de investigación mexicanos tienen mayor afinidad por videos de tipo infográficos, pues lo consideran un formato mucho más amigable para la asimilación de contenidos enfocados en la enseñanza.

Al inicio de la investigación, el equipo se había planteado la elaboración de infografías que permitieran a los usuarios revisar en una sola imagen, la mayor cantidad de información posible sobre la temática a abordar. Sin embargo, al someter los formatos a votación de un grupo de periodistas de investigación, se llegó a la conclusión de que los videos infográficos son mucho más didácticos y afines a las necesidades reales de los usuarios.

⁴⁴ Jakob Nielsen y Hoa Loranger (2007) Usabilidad. Prioridad en el diseño Web, ANAYA

6.7.2 Desarrollo de los contenidos

La fase 2 de la metodología que se aplicó al proyecto (observación participativa) no solo sirvió para profundizar en las necesidades y niveles de conocimientos de los periodistas de investigación, además facilitó la definición de contenidos a abordar en los diferentes módulos que conforman el sistema.

Desde un inicio, se había planteado que los contenidos a albergar en Dataísta, girarían alrededor de la metodología propuesta por el P.D. Sin embargo, hasta ese momento no se tenía nociones de hasta qué puntos se abordaría. Como resultado de la observación participativa se definió que los puntos a abordar dentro de Dataísta serían los siguientes:

■ **Introducción.-** Módulo inicial dentro del sistema en el cual se abordan conceptos básicos como base de datos, datos abiertos, P.D. Además, se incluye una breve explicación de cómo hacer uso del sistema y lo que el usuario puede encontrar en cada una de los módulos.

■ **Obtención.-** Módulo correspondiente a la Fase 1 de la metodología del P.D., en él se explica en qué consiste esta fase además de compartir con el usuario algunos tips para realizar búsquedas más eficientes para sus investigaciones.

■ **Scraping.-** Es el único submódulo dentro del sistema, se concibió de esta manera porque el *scraping* es una técnica que comúnmente se utiliza para ayudar a obtener información que enriquezca la investigación. En él se explica en qué consiste y se comparten algunas herramientas que pueden ser de utilidad para el periodista.

■ **Limpieza.-** Se explica en qué consiste, se dan algunos tips para poder realizar esta fase de la metodología de forma más sencilla, además de compartir algunas técnicas que ayudarán a simplificar esta labor al periodista.

■ **Análisis.-** Se explica en qué consiste esta fase, se comparten las formulas básicas más utilizadas para analizar datos. Además, enfatiza la importancia de no forzar a la base de datos para obtener los resultados esperados.

■ **Visualización.-** Último modulo en el que se comparte con el periodista, algunos conceptos básicos de diseño, que le permitirán seleccionar la mejor representación visual para su investigación.

Cuando se planteó el contenido de cada uno de los videos que explicarían la metodología,

se decidió manejar estructuras similares, es decir, se optó por abordar los puntos siguientes:

■ **¿Qué es?** - Concepto

■ **¿Por qué es importante?** - Relevancia dentro de la metodología

■ **¿Qué tomar en cuenta?** - Pasos a seguir a la hora de llevar a cabo la fase de la metodología

■ **Tips** - Con base en la información recopilada durante la observación participativa, así como de las entrevistas a los periodistas de datos, se identificaron algunas sugerencias o consejos que pueden ser de gran ayuda para aquellos que no tienen experiencia trabajando con bases de datos.

Se llegó a ellos como resultado del cruce de información recopilada en el “Taller de Datos Extremos” así como de la información contenida en los diversos manuales consultados durante la investigación documental. Además de permitir sintetizar y organizar la información más importante a comunicar durante los videos, la estructura, permite que las transiciones sean mucho más amigables para el usuario, pues todos los puntos que se abordan están relacionados entre sí.

Una vez que se definieron de manera clara los contenidos de cada módulo, el siguiente paso

fue realizar los *storyboards* correspondientes, en ellos se plantearon cada una de las animaciones además de establecer qué tipo de apoyos visuales se utilizarían.

Se optó por hacer uso de iconografía similar a la que se desarrolló para la identidad gráfica de Dataísta, para que hubiese consistencia entre los videos y el sistema. También, para que la narrativa visual fuese directa y al punto, es decir, que la animación no pesara más que lo que en el video se estaba comunicando. No se buscaba que la animación distrajera al usuario de lo que realmente importaba, que era entender el paso a paso de la metodología.

6.8 Descripción del Sistema de Información Dataísta

Al ser la categoría D.A. un concepto central en el desarrollo del trabajo de investigación era importante seleccionar un conjunto de tecnologías abiertas y bien documentadas por su comunidad de usuarios, para construir Dataísta. En primer lugar, al considerar Dataísta como un conjunto de recursos de información se decidió utilizar un software de gestión de contenido (CMS) de código abierto, para administrar los recursos de información. Así también, se seleccionó por su facilidad para diseñar e implementar en la web.

El software más cercano a las necesidades de construcción del sistema de información fue WordPress, por las siguientes características:

- Es software de código abierto.
- El uso de WordPress no requiere un pago. Solo el hospedaje del sitio tiene costo.
- Es la plataforma más utilizada para la publicación en línea, cerca del 20% de la web está construida en WordPress.
- Existen gran variedad de proyectos, los cuales comparten en código abierto su estructura y diseño.
- Tiene la posibilidad de crear diseños personalizados con características tecnológicas robustas.
- Tiene una amplia comunidad de usuarios y desarrolladores que pueden apoyar u orientar durante el proceso.

Las principales ventajas de este software de gestión de contenido son las siguientes:

- 1) El uso de diseños adaptables,
- 2) Diseños dentro del paradigma SEO,
- 3) El uso de plantillas responsivas, al menos en los diseños más generales,

- 4) Gran rendimiento en los procesos,
- 5) Administración rápida e inmediata,
- 6) Niveles óptimos de seguridad,
- 7) Herramientas poderosas para el tratamiento de multimedia y,
- 8) De fácil uso y acceso (WordPress, 2017).

Las características anteriores llevaron a tomar la decisión de utilizar WordPress como base para el diseño e implementación del sistema de información Dataísta. Otra de las ventajas de este gestor de contenido es la posibilidad de habilitar diferentes cuentas de administrador para

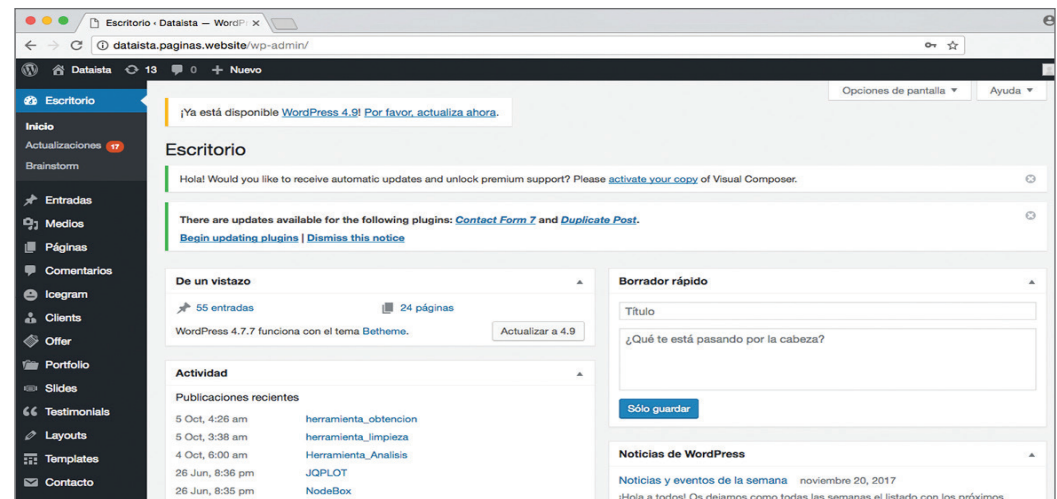


Figura 50. Editor de WordPress.

configurar las particularidades del sitio. Todo esto a través de un gestor gráfico que es capaz de integrar distintos elementos de información en una o más plantillas, además de proporcionar un editor de código fuente, por ejemplo HTML, CSS y JavaScript, por tratarse en este caso de un desarrollo Web. En la Figura 51 es posible observar el editor gráfico y en la Figura 52 el editor de código fuente.

Para la gestión de los videos se eligió el administrador de multimedia VIMEO, porque permite configurar una cuenta privada para compartirlos solo dentro del sitio del sistema de información. Se prefirió esta tecnología en vez del tradicional YouTube debido a que este no tiene opciones de seguridad, a menos de utilizar la versión de pago, para proteger a los recursos de su diseminación fuera del sistema de información.

La programación del sitio web del sistema de información se hizo, principalmente, con el editor visual del gestor de contenido. Por ejemplo, a partir de ello fue posible insertar las imágenes de cada una de las etapas, el logotipo, los botones, etc. Sin embargo, este gestor tiene limitaciones cuando se busca personalizar los estilos. Por ello, se utilizó el editor de código fuente para agregar los estilos CSS y agregar las acciones con código JavaScript. La parte interdisciplinaria del proyecto terminal ayudó en la estructuración de los textos instruccionales y en la elección de las imágenes adecuadas, tipografías y ubicación de los elementos de información y las secciones. Es importante mencionar lo anterior porque así, todos los elementos se pudieron ubicar en los mejores lugares para lograr una mejor comunicación de los mensajes emitidos.

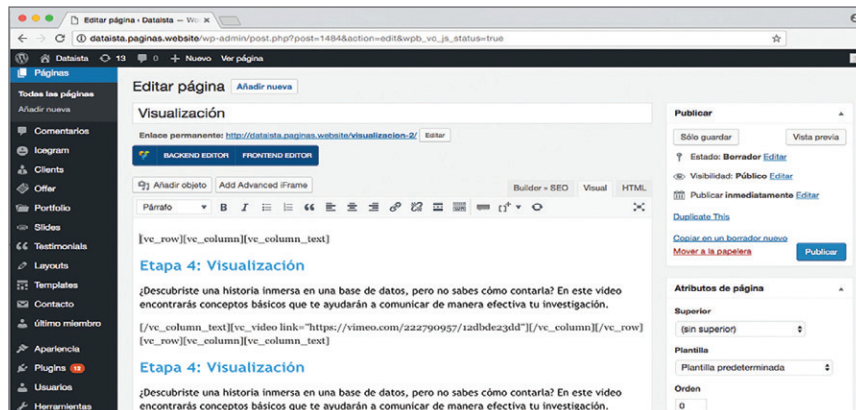


Figura 51. Editor Visual de WordPress.

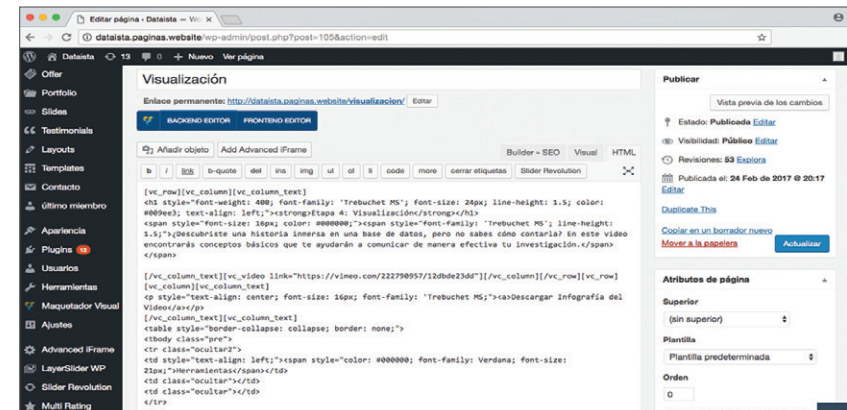


Figura 52. Editor de Código Fuente de WordPress.

6.8.1 Descripción general del prototipo

La esencia de Dataísta es proporcionar a los periodistas de investigación un sistema de información instruccional para facilitarles el entendimiento de las técnicas del P.D. Es decir, la obtención, la limpieza, el análisis y la visualización de bases de datos. Como se mencionó en el apartado número seis, de este mismo trabajo, el prototipo alberga información esencial para el entendimiento de las técnicas arriba mencionadas. Así mismo proporciona un vínculo entre periodistas, quienes, ya hacen P.D. y, aquellos interesados en incursionar en él.

Para ello, se desarrolló un prototipo de sistema de información con las siguientes características: estructurado en torno a un conjunto de recursos organizados e interrelacionados dinámicamente para proporcionar a los usuarios acceso a información relevante. En el caso específico de Dataísta los recursos organizados son: vídeos instruccionales, un sistema de clasificación de herramientas, foros, nube de *tags* o etiquetas, preguntas frecuentes, herramientas auxiliares, listado de bases de datos abiertos y casos de éxito P.D. alrededor del mundo (Heredero, y otros, 2006).

Todos los elementos contenidos en Dataísta están organizados por medio de secciones identificadas por cada una de las etapas del P.D.: obtención, limpieza, análisis y visualización.

En conjunto los recursos organizados y sus interrelaciones dinámicas buscan estructurar un sistema de información instruccional viable para satisfacer las necesidades de algunos periodistas de investigación mexicanos, interesados en obtener habilidades básicas en el manejo de datos.

Al combinar la organización de recursos de información y sus interrelaciones se puede definir a Dataísta como un sistema de información, donde, a través de imágenes, videos y texto se reflejan las necesidades de formación y aprendizaje en el uso de las técnicas del P.D. (obtención, limpieza, análisis y visualización) de los periodistas de investigación mexicanos. Dataísta busca habilitar tecnológicamente en el uso de datos a los periodistas de investigación mexicanos con el fin de que puedan obtener un mejor rendimiento en el uso de los datos abiertos para contar historias periodísticas y fortalecer la transparencia y la rendición de cuentas.

6.8.2 Página principal

Está integrada por el logo, una leyenda y un vídeo de bienvenida, los cuales buscan dar al usuario de Dataísta, información precisa para entender de forma general los fines y el alcance del prototipo. Asimismo, desde esta pantalla es posible acceder al contenido del sistema de información a través del registro de usuario

por medio de la opción de inicio de sesión o sin registro a través de un botón azul debajo del video de bienvenida.

Un usuario registrado será capaz de realizar las siguientes actividades: **1)** comentar en los foros, **2)** utilizar el clasificador de herramientas, **3)** utilizar la nube de etiquetas para buscar contenidos, **4)** sugerir la incorporación de bases de datos al catálogo y **5)** retroalimentar las bases de datos a partir de sus experiencias de uso. Un usuario no registrado solo tendrá permitido: **1)** navegar en las secciones de la metodología, **2)** observar la clasificación de herramientas, **3)** leer los foros, **4)** visualizar las principales palabras de la nube de *tags* o etiquetas y no podrá utilizar las opciones de sugerencia o retroalimentación al catálogo de bases de datos abiertos.

Una parte importante de esta sección son los botones de la parte de arriba que indican medios, datos abiertos, preguntas frecuentes y el inicio de sesión. Pues cada una de ellas lleva a una sección de contenido del sistema de información, pero también aparece en todas las secciones y contenidos del sistema como un medio para facilitar la navegación entre ellas. El logo funciona también como un enlace. Si se es un usuario no registrado al presionar sobre el logo será direccionado a la misma página principal, por otra parte, si se es un usuario registrado este será llevado a la sección de *Home*, es decir a la página principal.

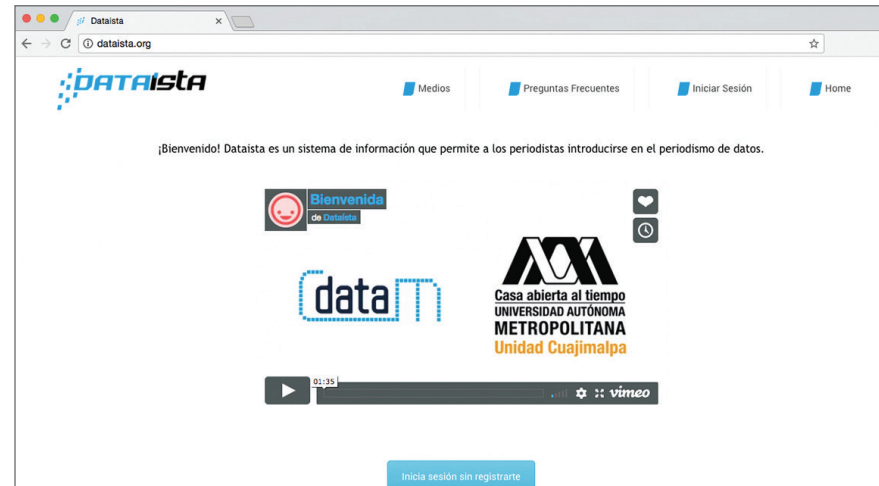


Figura 53. Página de Bienvenida del Sistema de Información, Dataísta.

6.8.3 Home

En esta sección existe un menú formado por los iconos representativos de cada una de las etapas de la metodología del P.D., así como por la introducción. Al posicionar el cursor sobre cada una de las imágenes es posible ver una breve descripción de cada etapa. Este apartado está acompañado de un texto instruccional, el cual invita a explorar cada una de las etapas. A partir de esta página es posible empezar a navegar dentro de las secciones y recursos de información del sistema Dataísta. Al igual que en todas las secciones del sistema esta página utiliza el menú de la parte superior para permitir la navegación con las secciones principales del sistema.

6.8.4 Introducción

Se diseñó una sección de introducción que sirva como una entrada para el usuario en los conceptos básicos, en la figura número 55 se muestra la página de Introducción al contenido en el sitio.

6.8.5 Secciones de Obtención, Limpieza, Análisis y Visualización

Para estas secciones se diseñó una plantilla de contenido, pues esta es la parte fuerte del proyecto a propósito de cuestiones instruccionales. Las características de esta plantilla son: un texto instruccional, un video, una tabla con una clasificación de herramientas y una herramienta auxiliar. Así como un menú lateral de lado izquierdo para navegar entre cada una de las fases, el cual es fijo.

Otra característica son dos elementos ubicados a la derecha de la página, donde se ubican enlaces para acceder a cada uno de los foros de discusión y una nube de palabras donde se observan las entradas de contenido más utilizadas. Las entradas indexadas en la nube pertenecen a la descripción de herramientas y los foros, esta característica pretende mostrar los temas más populares a los que acceden los usuarios. Los cuatro apartados

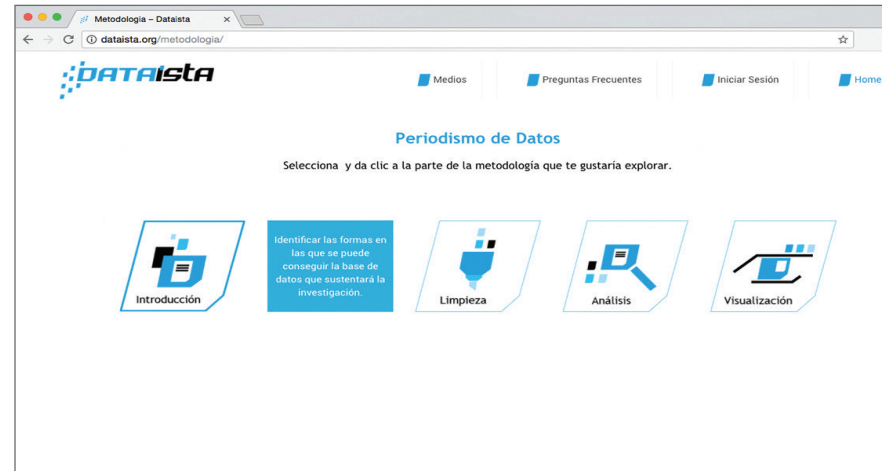


Figura 54. Home del Sistema de información, Dataísta.



Figura 55. Página de Introducción al contenido.

en donde se describe cada una de las fases de la metodología del P.D. siguen esta misma disposición, sin embargo, en algunos casos cambian en cuestión de diseño e información.

Al igual que en otras secciones la barra de opciones superior se mantiene fija. Es preciso mencionar la utilidad de estas áreas de opciones fijas, pues son de utilidad para la navegabilidad en el sitio. Esto se diseñó con la idea de proporcionar al usuario una experiencia de uso integral, es decir contar con la posibilidad de moverse entre secciones del contenido.



Figura 56. Sección de Texto Instruccional y Video.

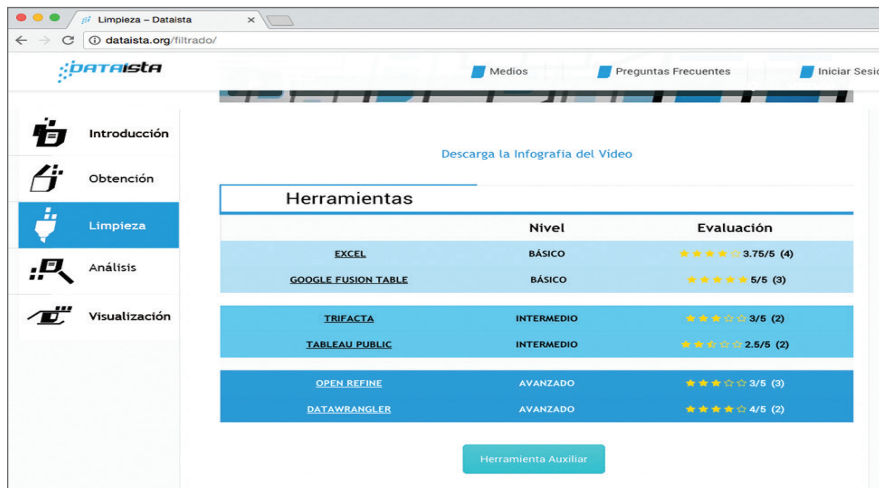


Figura 57. Sección de Clasificación de Herramientas.



Figura 58. Sección de Herramienta Auxiliar.

6.8.6 Medios

En esta sección se presenta de manera infográfica los principales proyectos de P.D. en el mundo, pero principalmente en Latinoamérica y en México. Esta parte tiene la finalidad de proporcionar al usuario un recorrido sobre las características de estos proyectos para así poder recorrer las demás secciones de contenido con un mayor conocimiento de este tipo de iniciativas.

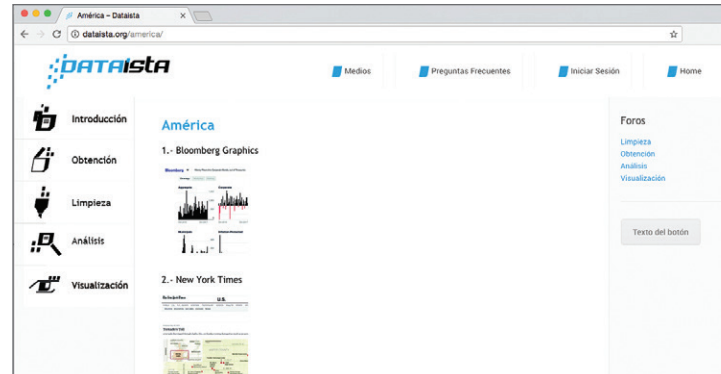


Figura 59. Sección de Medios.



Figura 60. Sección de Preguntas Frecuentes.

6.8.7 Preguntas Frecuentes

Esta sección pretende mostrar al usuario información precisa sobre el objetivo y finalidad del sistema de información. Esta parte es clave para orientar a todos los usuarios acerca de lo que pueden hacer y no hacer con Dataísta. Son 10 las preguntas con su respectiva respuesta, las cuales se refieren al impacto social del sistema, algunos asuntos técnicos o el desarrollo del proyecto. Esta sección se complementa en contenido con el video de la página principal, donde se explica qué es Dataísta.

6.8.8 Foros

Los foros son un medio de comunicación entre los usuarios registrados al sistema de información. Existen 4 foros, cada uno corresponde a una etapa de la metodología del P.D. Al hacer clic en cada uno de ellos el usuario será llevado a los comentarios y discusiones de cada tema.

Es preciso señalar que cada usuario puede publicar lo que desee, sin embargo, el administrador del sistema podrá elegir qué comentarios serán permitidos y visualizados y cuáles no. Esto con la finalidad de evitar comentarios fuera de lugar. Los usuarios no registrados podrán leer la sección de foros, pero no podrán participar en ellos.

Los foros resultan una parte muy importante en el sistema debido a posibilidad de integrar a los usuarios a una comunidad todavía pequeña. Así, como en muchos otros sitios, el foro se vuelve un punto importante de reunión permitiendo que sean ellos mismos los que den respuesta a las inquietudes de la comunidad.

6.8.9 Inicio de Sesión o Registro

Esta sección se utiliza para acceder al contenido del sistema como usuario registrado. Así como también al panel de administrador para realizar cambios en el sistema de información.

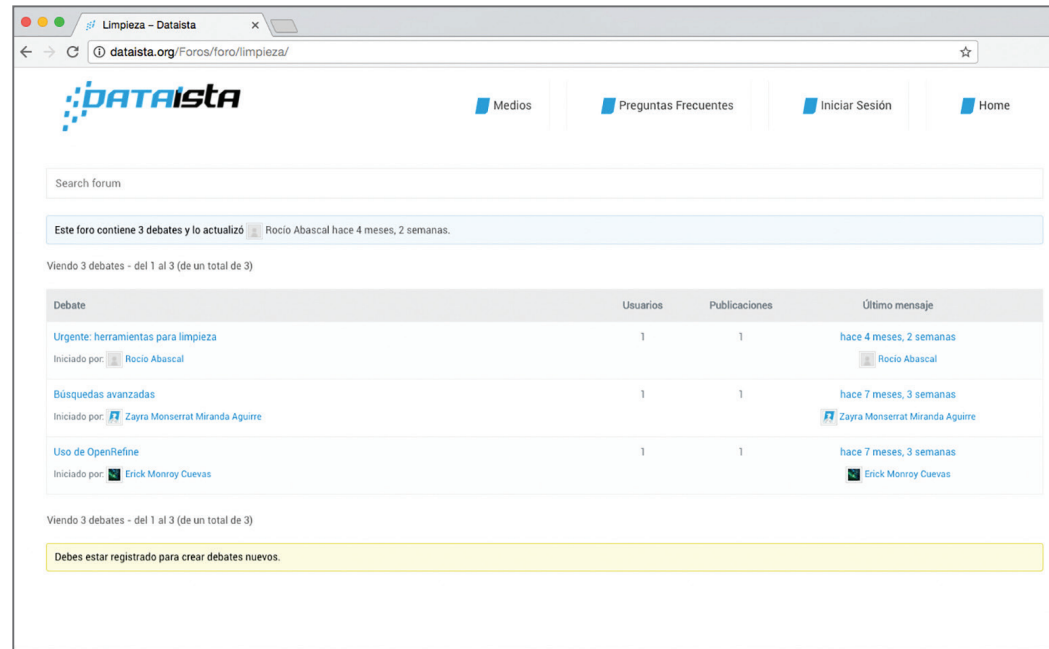


Figura 61. Sección de Foros.

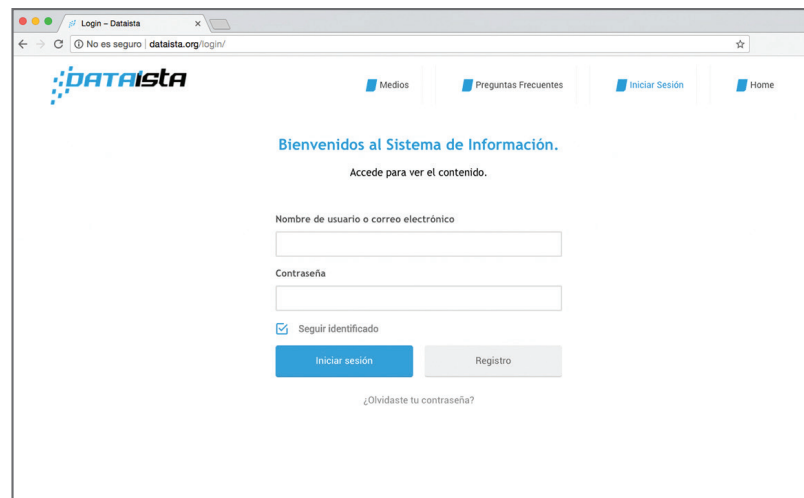


Figura 62. Sección de Inicio de Sesión y Registro.

6.8.10 Datos Abiertos

En este apartado se enlistan diferentes Bases de Datos Abiertos con la finalidad de orientar a los usuarios en la búsqueda de conjuntos de datos sobre diversos temas. Se clasificaron los conjuntos de datos en 4 categorías: México, internacionales, organizaciones no gubernamentales y académicas.



Figura 63. Sección de Datos Abiertos.

6.8.11 Herramienta Auxiliar

La sección de herramientas auxiliares proporciona soluciones sencillas a algunos problemas típicos del tratamiento de datos. Se puede acceder a ellas a partir de un botón debajo de la tabla clasificadora de herramientas en cada etapa de la metodología.

Las herramientas auxiliares son 4, la finalidad es que sean integrales y cada una de ellas sea la continuación de la anterior. Esto es posible en el caso de limpieza, análisis y visualización, pues la base de datos con la que se trabajó puede pasar por cada una de las etapas anteriores. En el caso de la herramienta de obtención es posible utilizar los documentos generados por ella, sin embargo, en el trabajo del periodista de datos, frecuentemente, se utilizan documentos de Excel (xls) que son creados de manera manual. Las herramientas auxiliares son las siguientes:

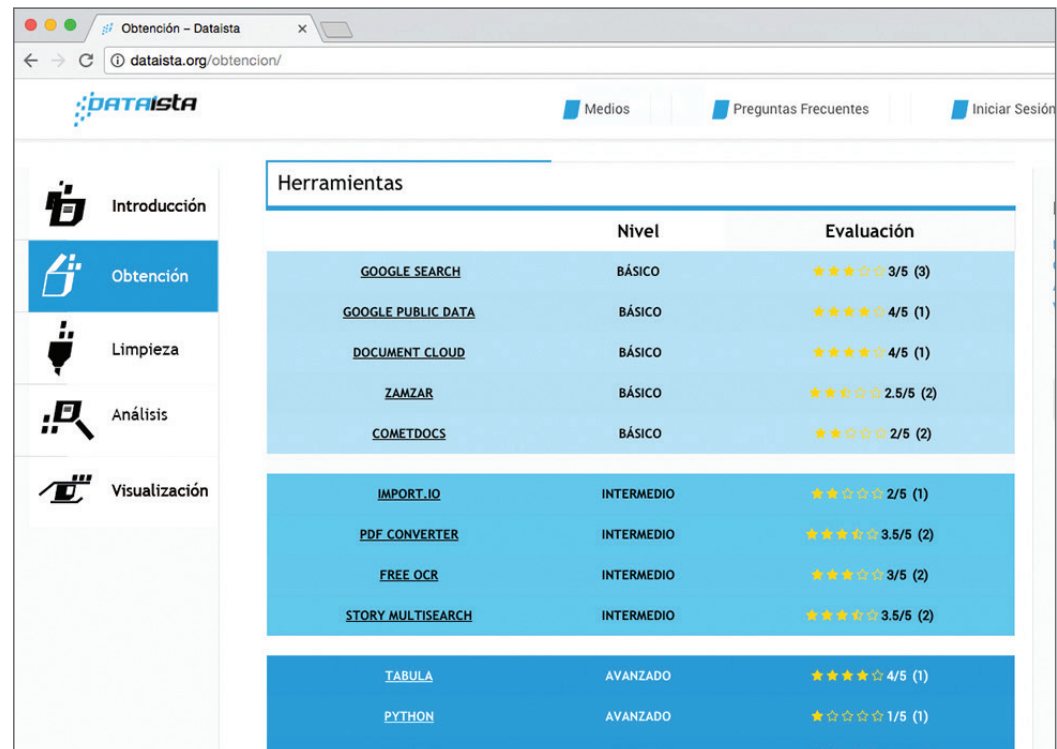


Figura 64. Vista del Clasificador de Herramientas.

6.8.12 Clasificador de Herramientas

Esta sección tiene como finalidad mostrar una clasificación de las principales herramientas utilizadas por los periodistas de datos. Se estructuraron con base en tres niveles de dificultad; básico, intermedio y avanzado. Para cada nivel se asignó un color, en este caso tonos de azul, para diferenciar visualmente cada una de ellas.

Al hacer clic en alguna herramienta se puede observar el desglose de ella, sus datos principales: nombre, qué se necesita para usarse y posibles aplicaciones. También es posible desde aquí evaluar cada herramienta para que aparezca en el clasificador en una escala de 5 representada por estrellas, como puede observarse en las figuras anteriores.

6.9 Desarrollo de las Herramientas Auxiliares

En el periodismo de datos se utiliza el poder de las computadoras para descubrir, contrastar, visualizar y combinar información de varias fuentes. El dato utilizado por esta tendencia, del periodismo de investigación, tiene una naturaleza específica: la posibilidad de ser procesado por sistemas informáticos y potentes herramientas de computación (Flores Vivar, 2012).

El Periodismo de Datos es la más conocida de

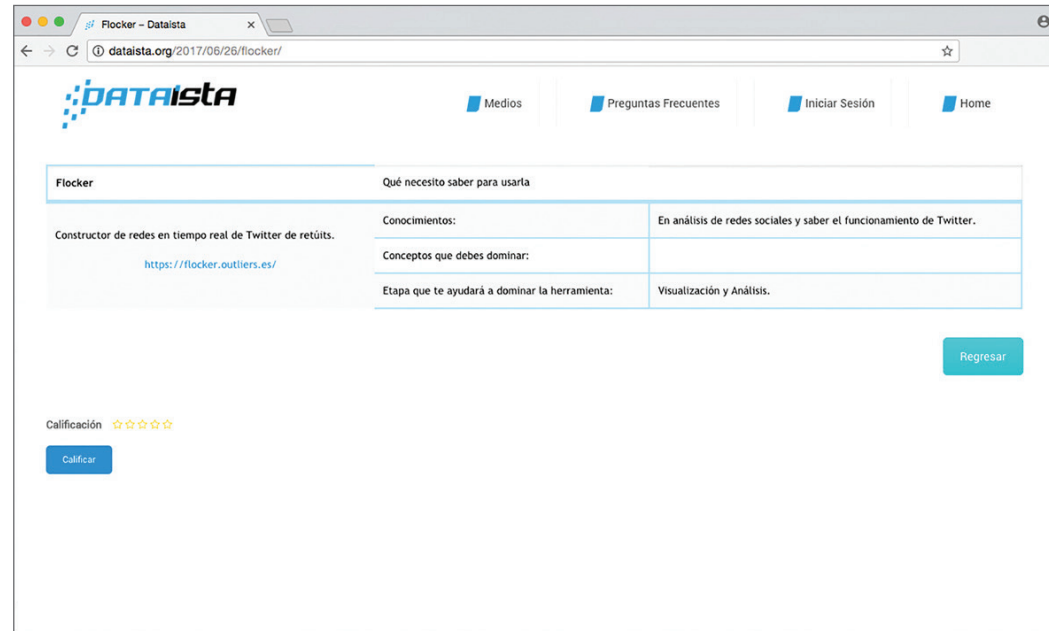


Figura 65. Detalle del Clasificador de Herramientas.

las nuevas tendencias en el uso de la tecnología digital en el periodismo, pero no la única. Pues, también se pueden mencionar a los juegos de noticias, el periodismo de drones y el de realidad virtual como iniciativas exploratorias en esta combinación. De manera particular, el P.D. está evolucionando hacia la aplicación del aprendizaje automático y los algoritmos para contar historias con valor periodístico. Sus principales áreas de trabajo son la información no estructurada como: texto, video y audio. Esto como una clara oposición a los datos históricamente

utilizados por los periodistas: hojas de cálculo y bases de datos (Charles & Cheryl, 2016).

El profesional del P.D. debe tener noción de los siguientes elementos: los procesos de digitalización, la filosofía de los datos abiertos, la transparencia, el acceso público a la información, nociones estadísticas, herramientas de visualización y habilidades para la investigación. La incorporación de estas habilidades a la práctica periodística permite transformar meros datos en una historia atractiva, a partir de la compilación, el filtrado, la contextualización, el contraste, la organización y la jerarquización (Rodríguez Brito & García Chico, 2013).

El P.D. se divide en cuatro categorías (ver anexo A), las cuales en la práctica se interrelacionan de manera inexorable. La primera es el reporte de datos (*data reporting*) el cual consiste en obtener, limpiar y analizar datos para contar historias periodísticas. La segunda es la visualización de datos e interactivos, donde, se utiliza código para la publicación (HTML, CSS, JS) y bases de datos para construir trabajos periodísticos interactivos. La tercera es la emergencia de tecnologías para el periodismo, la cual consiste en desarrollar herramientas y productos usando datos y tecnología. La cuarta es el periodismo computacional, donde se utilizan algoritmos, el aprendizaje automático y otros métodos para apoyar a los objetivos del periodismo (Charles & Cheryl, 2016).

El papel de los programadores y la programación en el periodismo de datos consiste en la

creación de secuencias de comandos para facilitar a los periodistas y a los usuarios el manejo de tareas complejas en materia de programación. Su importancia radica en la posibilidad de crear estas secuencias para hacer más fácil a los periodistas y a los usuarios el poder obtener y aprovechar bases de datos y recursos de información. En este sentido, el sistema de información Dataísta tiene como objetivo ayudar a los periodistas de investigación mexicanos en el aprendizaje de la metodología del P.D., para obtener, explorar, limpiar, analizar y visualizar de manera sencilla patrones y tendencias dentro de las bases de datos (Flores Vivar, 2012).

El periódico británico *"The Guardian"* sugiere a los programadores y desarrolladores de software para el P.D., la construcción de herramientas intuitivas, de fácil uso para cualquier usuario y con la capacidad de utilizarse en diferentes plataformas. Así como también, liberar a los periodistas del manejo directo de lenguajes de programación o de inscribirse en costosos y engorrosos cursos de actualización, tanto presenciales como en línea (Bradshaw & Rohumaa, 2011).

Para el diseño de las herramientas auxiliares se eligió como lenguaje de programación Python. Debido a que muchos proyectos de manejo de datos utilizan este lenguaje, es fácil encontrar en la web y la literatura especializada numerosos ejemplos y aplicaciones para la obtención, limpieza, análisis y visualización de datos. Esto permite, también, contar con una amplia comunidad de usuarios y desarrolladores.

Python es un lenguaje de programación de fácil lectura, gracias a su sintaxis limpia y a la facilidad para invocar librerías y clases. Sin embargo, para su uso en ambientes web resulta algo engorroso. Debido a ello, se siguieron tres fases en el diseño de las herramientas:

- 1) Programación de la herramienta para su uso local por medio de la terminal,
- 2) Uso de una interfaz gráfica para probar la usabilidad de la herramienta,
- 3) Adecuación a un ambiente web. En el apartado de las herramientas se profundizará más sobre este punto.

El objetivo principal de los esfuerzos por acercar a los periodistas hacia la programación consiste en la construcción de un lenguaje en común. Esto para poder identificar las áreas de oportunidad y las posibilidades reales para convertir datos en historias periodísticas. En ningún momento se pretende convertir a los profesionales del periodismo en programadores expertos (Bradshaw & Rohumaa, 2011).

Una sección del sistema de información, Dataísta, es la de herramientas para cada una de las etapas de la metodología del periodismo de datos. Para el diseño de estas se utiliza el paradigma de los sistemas concisos para la creatividad computacional. Esta perspectiva consiste en el diseño de pequeños programas informáticos

capaces de reflejar de manera sencilla, rápida, concisa y a partir de poco código algunos procesos o procedimientos de la creatividad a partir del uso de modelos computacionales (Montfort, 2015).

Esta noción de los sistemas concisos puede traslaparse al P.D., a partir del diseño e implementación de sencillas herramientas para mostrar a los periodistas cómo trabajar con datos en cualquier etapa de la metodología (obtención, limpieza, análisis y visualización). Lo anterior se justifica a partir de las conclusiones de un estudio realizado por Fontichiaro (2016) donde se logró identificar la importancia de la alfabetización en datos como una habilidad fundamental para comunicar información de una forma poderosa a partir de visualizaciones. Sin embargo, el estudio analizó la desconexión entre las habilidades desarrolladas por alumnos de materias como: estadística, diseño y computación encontrando que están desconectadas de las necesidades del mundo real.

Con base en la recolección de datos cualitativos, a través de entrevistas y observación participativa, se encontraron algunos hallazgos sobre la manera en cómo los periodistas de datos mexicanos han afrontado los retos inherentes a las tareas de limpiar, analizar y obtener datos. Entre ellos los más relevantes corresponden a la etapa de análisis de información, en la cual se observó la desconexión mencionada por Fontichiaro. Esta desconexión consiste en la carencia, por parte de los periodistas, de un

conocimiento previo para el manejo de datos, lo cual, dificulta la apropiación de las sofisticadas y potentes herramientas para el manejo de datos existentes en el mercado.

Un estudio realizado por Bhargava y D'Ignazio (2016) señala la proliferación de herramientas centradas en la resolución de problemas específicos en el manejo de datos. Principalmente, para ayudar a los usuarios a recolectar, trabajar y visualizar datos. Para estos investigadores esta perspectiva es poco adecuada, si la finalidad es alfabetizar a los usuarios en el manejo de datos, pues no fortalecen las habilidades y conocimientos para hacer un uso fluido de los datos.

Por ello, se utiliza el paradigma de los sistemas concisos para diseñar herramientas para el manejo de datos por parte de los periodistas de investigación mexicanos quienes buscan acceder a la realización de investigaciones periodísticas a partir de datos. Con base en los principios desarrollados por Bhargava y D'Ignazio (2015) para el diseño de herramientas para el manejo de datos con un enfoque pedagógico, se busca que las herramientas de Dataísta permitan al usuario: **1)** enfocarse en el desarrollo de una sola tarea, **2)** guiarlo a través del proceso para aprender el razonamiento detrás de ella, **3)** que las herramientas sean amigables y lúdicas para invitar al usuario al aprendizaje y **4)** la capacidad de trabajar en conjunto entre herramientas.

Otro enfoque importante para el diseño de las herramientas del sistema de información Dataísta es el paradigma de la programación exploratoria.

Este consiste en la posibilidad de diseñar programas informáticos claros y sencillos para que usuarios no expertos sean capaces de comprender la lógica con la que se resuelve un problema a partir de una pieza de software (Montfort, 2016).

En suma, las herramientas de Dataísta están construidas con base en las siguientes características: concisas, sencillas, con poco código para dar al usuario la posibilidad de entender la lógica detrás del código, centradas en el aprendizaje de cada una de las etapas de la metodología del P.D., orientadas en el aprendizaje de un procedimiento más que en la resolución de un problema específico y con la posibilidad de integrar más herramientas.

El lenguaje de programación con el que se construyeron las herramientas corresponde a Python y sus principales librerías para el tratamiento de datos como: *BeautifulSoup*, *Pandas*, *Pygal* y *Matplotlib*. Así como la librería *Flask* para adaptar los programas en Python al entorno de la web. En las siguientes secciones se describe cada una de estas herramientas de desarrollo.

6.9.1 Python

Python es un lenguaje de programación interpretado orientado a objetos, disponible en múltiples plataformas y sistemas operativos. Su principal fortaleza radica en ser un estándar de programación, pues cualquier *script* de Python puede ser utilizado, modificado y probado en cualquier

sistema con el intérprete instalado, ya sea Windows, Macintosh o Unix (Telles, 2008).

Python ha sido utilizado para desarrollar sistemas de uso local en una sola computadora, así como aplicaciones basadas en la web para correrse en todo el universo de las computadoras. Permite construir aplicaciones para el área de los negocios, las ciencias, los videojuegos, así como la automatización de algunas rutinas de actualización en Internet para bancos y otras instituciones financieras (Telles, 2008).

El periodista Cheryl Phillips considera a Python como un lenguaje de programación muy útil para el desarrollo de proyectos de P.D. Entre las razones para ello, se pueden mencionar su estatus de código abierto, la facilidad para escribir y leer código, reglas sintácticas bastante simples y flexibles. Pero la característica más sólida de Python es la existencia de una inmensa comunidad activa de usuarios y la disponibilidad de librerías para el tratamiento de datos como: Pandas, BeautifulSoup, Numpy, Matplotlib, Pygal y Bokeh (Bradshaw & Rohumaa, 2011).

Ben Welsh editor en “Los Angeles Times Data Desk” utiliza Python, junto con un equipo de P.D., para recolectar, organizar, analizar y presentar grandes conjuntos de datos, principalmente la librería Pandas. En julio de 2017 este equipo junto con el “Knight Center for the Journalism in the Americas” dictaron un curso masivo en línea llamado “Python para Periodistas de Datos: analizando el dinero en la política”. Su principal objetivo fue utilizar Pandas para leer, filtrar, unir,

agrupar, agregar y clasificar datos estructurados del “California Civic Data Coalition” (Welsh, 2017).

Otro proyecto de P.D. utilizando Python es el taller “Coding for Journalism” realizado en la conferencia del “Investigative Reporters and Editors” del año 2015. Actualmente se ofrece como un tutorial en la web cuyo principal objetivo es introducir a los periodistas de manera paulatina al desarrollo de aplicaciones web utilizando el marco Flask, otra librería de Python, por medio de nueve prácticas de *scraping* de páginas web y limpieza de bases de datos (Investigative Reporters and Editors, 2015).

Python es un lenguaje de programación versátil pues, así como es posible escribir *scripts* de análisis de datos, también se pueden desarrollar aplicaciones web. Entre sus principales características están: su extensibilidad, ser software libre, su compactibilidad y ser un estándar. Las principales librerías para el desarrollo del P.D. son: BeautifulSoup para la obtención de páginas web, Pandas para el análisis de tablas y series de datos, Pygal y Matplotlib para la visualización de datos en forma de gráficos, interactivos el primer caso y estáticos en el segundo (Kara, 2016).

6.9.2 Librería Flask

Flask es una librería de Python para el desarrollo de aplicaciones. Está basado en la especificación interfaz de puerta de enlace de un servidor Web (WSGI) de Werkzeug y el motor de plantillas

Jinja2 bajo una licencia Berkeley Software Distribution⁴⁵. Se considera micro porque su núcleo es pequeño, pero altamente extensible para el desarrollo de aplicaciones web. Permite a los desarrolladores elegir la configuración deseada sin restricciones de bases de datos o plantillas (Aggarwal, 2014).

Werkzeug es una biblioteca WSG para Python. Es ampliamente utilizado bajo una licencia Berkeley Software Distribution. Comenzó como una colección simple de varias utilidades para aplicaciones WSGI y se ha convertido en uno de los módulos de utilidad WSGI más avanzados. Incluye un poderoso depurador, objetos de respuesta y solicitud completos, utilidades HTTP para manejar etiquetas de entidad, encabezados de control de caché, fechas HTTP, manejo de cookies, carga de archivos y un potente sistema de enrutamiento de URL (Ronacher, Werkzeug, 2014).

Jinja2 es un motor de plantillas para Python. Cuenta con soporte Unicode y un entorno de ejecución integrado opcional, ampliamente utilizado y licenciado mediante Berkeley Software Distribution. Está inspirado en el sistema de plantillas de Django, sin embargo, se extiende con un lenguaje expresivo que proporciona a los desarrolladores de plantillas un conjunto de herramientas más potente. Además, integra la ejecución a partir de una caja de seguridad y escape opcional para aplicaciones. Entre las principales características de Jinja 2 pueden mencionarse las siguientes: cada aspecto de la ejecución de la plantilla es monitoreado explícitamente, cuenta con un poderoso sistema automático de

escape HTML para la prevención de secuencias de comandos entre sitios, permite la herencia de plantillas con lo que se puede utilizar el mismo o un diseño similar para todas las plantillas; modo de depuración con un sistema que integra la plantilla de compilación y errores de tiempo de ejecución en el sistema de rastreo estándar de Python (Ronacher, Jinja, 2014).

En términos simples un marco de desarrollo web es una colección de librerías cuya funcionalidad consiste en resolver una parte de los problemas genéricos que surgen al desarrollar cualquier aplicación. Por ejemplo, la asignación de las rutas URL hacia recursos específicos, el uso de contenido dinámico en HTML o la interacción con el usuario final. Es decir, solo las funciones esenciales son instaladas con la librería, mientras otras funciones más avanzadas pueden utilizarse a través de extensiones (Dwyer, 2016).

El código fuente de los proyectos web que utilizan Flask son fácilmente entendibles, pues con pocas líneas de código es posible adaptar poderosas soluciones. Se considera una solución bastante Phytonica⁴⁶ porque su curva de aprendizaje para los principiantes es menor, ya que se requiere poco código para conseguir una aplicación sencilla y funcional. Al igual que Python este marco web cuenta con un amplio espectro de desarrolladores, tutoriales y materiales de aprendizaje (Ronacher, Flask Web Development, 2017).

Usualmente los programas de Python se muestran en la línea de comandos de la computadora. Sin embargo, a medida que una aplicación se vuelve más compleja es necesario convertirla

⁴⁵ La BSD es una licencia permisiva sin copyleft. Consta de tres partes: leyenda copyright, licencia y limitación de responsabilidad.

⁴⁶ Es una característica deseable del código escrito con lenguaje Python, el cual es legible y transparente. Para mayor información consultar el documento el Zen de Python.

en aplicaciones de escritorio, donde el ratón y el teclado juegan un papel fundamental para la entrada de datos a través de menús y elementos interactivos. El siguiente paso es compartir con una audiencia mayor las aplicaciones para ello se utilizan los navegadores web (Dwyer, 2017).

Una de las ventajas de una aplicación web es su flexibilidad, pues se puede utilizar en distintos sistemas. Además, su condición de existencia solo en línea permite accederse a ella por medio de una conexión a Internet. Por otra parte, una aplicación web no tiene que ser descargada, instalada o reiniciarse (Meyer, 2011).

Para construir las herramientas auxiliares del sistema de información Dataísta se siguió un método compuesto por cuatro partes. La primera, fue un análisis de las principales características de las librerías de Python para el tratamiento de datos. La segunda, el diseño de las funciones básicas de los programas. En la tercera, se codificaron los *scripts*, primero en línea de comandos y a nivel local. En la cuarta, ya con los *scripts* codificados se integraron al flujo de trabajo del marco web Flask.

En los siguientes apartados se describe cada

una de las herramientas auxiliares. Primero, se explica brevemente la etapa para la que se desarrolló cada una de ellas. En segundo lugar, las técnicas utilizadas y las librerías. También se muestran algunas pantallas de las interfaces, Por último, en la sección de anexos se puede consultar el código fuente de cada una de las herramientas.

6.9.3 Herramienta Obtención

Esta herramienta permite introducir al periodista en algunos conceptos del *scraping* de datos. Específicamente, la recuperación de datos de una página web a partir de algunas etiquetas HTML. También permite extraer información de algunos documentos PDF. Por otra parte, esta diseñada para hacer posible la recuperación de datos de algunas páginas web. Los datos que se pueden recuperar son párrafos, tablas, enlaces y encabezados.

Figura 66. Etapas de Desarrollo de las Herramientas Auxiliares.



No.	Scraping Web	Extracción PDF
1	Obtener URL	Subir archivo
2	Elegir etiqueta a buscar: <p>, <table>, <a> y <h1>	Elegir de la bola de la que se obtendrá información
3	Descargar archivo txt con los datos de la etiqueta buscada	Descargar archivo txt con los datos recuperados

Tabla 7. Procesos de la Herramienta de Obtención.

Es importante mencionar que el *scraping* de datos web es una tarea muy particular y cada proyecto de P.D. tendrá requerimientos particulares, por ello es complicado diseñar una herramienta de carácter genérico. En el caso de la herramienta obten se busca mostrar a los periodistas algunas posibilidades de Python y BeautifulSoup para extraer información de la web. En el caso de la extracción de información de un documento PDF se utiliza la biblioteca PyPDF2, la cual proporciona algunas tareas básicas para extraer datos.

En esta herramienta se pueden realizar 2 procesos, el primero es el *scraping* de páginas web a partir de insertar una URL en la herramienta. Una vez con la URL recuperada, la herramienta da la posibilidad de obtener las etiquetas HTML de <p>, <table>, <a> y <h1>. Posteriormente, se puede descargar un archivo de texto (.txt) con el resultado del *scraping*. La tabla siguiente muestra de manera esquemática estos procesos.

Para Lawson (2015) el término *scraping* se refiere al proceso de extraer datos de una página web.

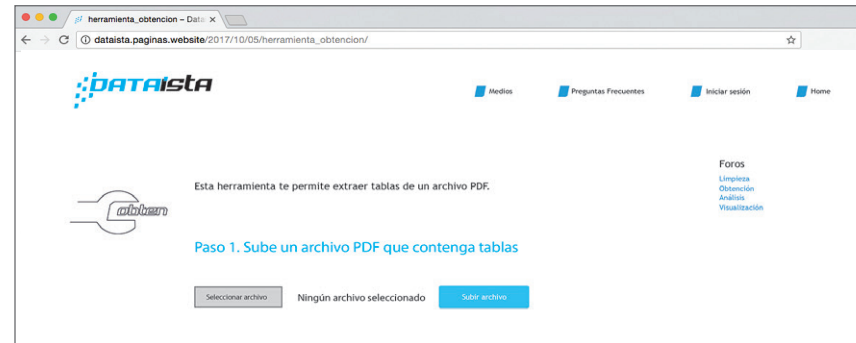


Figura 67. Interfaz Herramienta Obtención.

Es una técnica muy útil para enfrentarse a los problemas de reusabilidad y compatibilidad de formatos. El *scraping* se considera un arte por medio del cual se controla de forma automatizada sitios web con la finalidad de extraer datos publicados de interés (Carvajal, 2016).

Los principales tipos de herramientas para *scraping* pueden caer en alguna de las siguientes categorías de uso: recuperar etiquetas HTML a través del nombre del dominio, analizar la información de las etiquetas recuperadas, almacenar esta información para su posterior uso y repetir este proceso en otras páginas (Mytchell, 2015).

Un modelo característico del *scraping* es el denominado esquema “Orquestador + Navegador”. Este consiste en el uso de un programa, en cualquier lenguaje de programación, para controlar un navegador web. El orquestador es un programa de alto nivel, el cual envía comandos al explorador web para guiar la navegación en el sentido deseado. Por su parte, el navegador es una especie de terminal tonta, donde, se manejan los elementos de una página web, capaz de seguir las instrucciones del orquestador (Carvajal, 2016).

En resumen, el *scraping* es un conjunto de técnicas automáticas para la obtención de datos provenientes de páginas web a partir de código HTML. Estas técnicas se utilizan principalmente para recopilar grandes cantidades de datos para su homogenización, tratamiento y extracción de conocimiento (Moya, 2015).

Para la construcción de esta herramienta se utilizó principalmente la librería BeautifulSoup de Python para el manejo de datos en formato HTML y XML provenientes de la web. Proporciona un analizador capaz de navegar, buscar y modificar un árbol de análisis. Es una opción popular entre los programadores, debido a que permite ahorrar días y horas de trabajo. Fue creada en 2004 por Leonard Richardson un ingeniero en computación de la UCLA.

Esta librería transforma un archivo HTML o XML en un árbol de etiquetas a partir del cual es posible identificar, modificar y buscar etiquetas, clases y atributos, así como, recuperar los datos dentro de las propias etiquetas. Entre sus principales

características podemos mencionar: la navegación a través de la estructura de datos, extracción de URLs, la recuperación de texto y el uso de distintos analizadores de código HTML y XML.

Esta librería se usa principalmente para el *scraping* teniendo, tres poderosas características:

1. Proporciona métodos simples y *pythonicos* para navegar, buscar y modificar un árbol de análisis. Se considera una caja de herramientas para diseccionar un documento y extraer la información necesaria. No se necesita mucho código para escribir una aplicación.
2. Convierte automáticamente los documentos entrantes a Unicode y los documentos salientes a UTF-8. Esto permite dejar de pensar en codificaciones a menos que el documento no especifique una y la librería no sea capaz de detectar una codificación.
3. Esta librería está basada en los analizadores de Python como lxml y html5lib, lo que permite utilizar diferentes estrategias de análisis para evaluar velocidad y flexibilidad.

Para utilizar esta librería es necesario contar con conocimientos mínimos de código HTML y CSS. Aunado a lo anterior es preciso analizar la estructura HTML de la página a la cual se aplicarán las técnicas de obtención de datos, ello con la finalidad de identificar el contenido a recuperar.

6.9.4 Herramienta Limpieza

Esta herramienta permite cambiar o eliminar datos en un documento xlsx. También convertir un xlsx de mayúsculas a minúsculas y viceversa. Las librerías utilizadas en su construcción son: Pandas y las nativas de Python. La tabla siguiente muestra de manera esquemática estos procesos.

La Figura siguiente muestra la interfaz gráfica de la herramienta de limpieza:

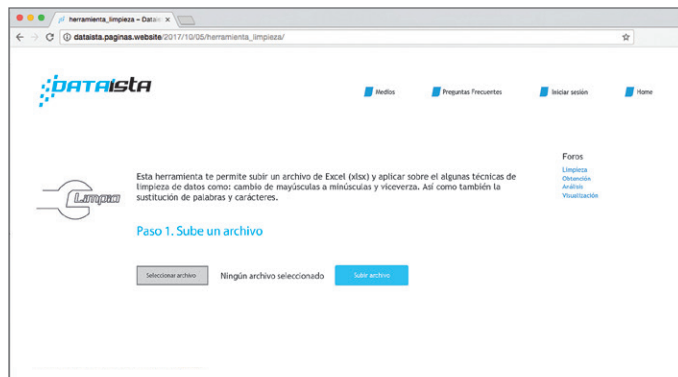


Figura 68. Interfaz de la Herramienta de Limpieza.

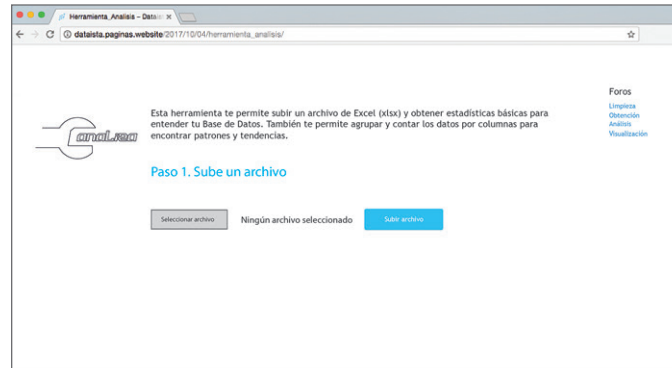


Figura 69. Interfaz de la Herramienta de Análisis.

6.9.5 Herramienta Análisis

Esta herramienta permite subir un archivo xlsx y obtener datos estadísticos sobre el cómo: la media, la moda y la mediana. También permite agrupar y contar los datos por columnas para encontrar algunos patrones. La tabla siguiente muestra de manera esquemática estos procesos.

Para construir esta herramienta se utilizó la librería de código abierto para el tratamiento de

Pasos	Limpieza de un archivo XLSX
1	Subir archivo
2	Elegir opción de limpieza: Sustituir palabras, minúsculas o mayúsculas
3	Descargar archivo limpio

Tabla 8. Procesos de la Herramienta de Limpieza

Pasos	Análisis de un archivo XLSX
1	Subir archivo
2	Elegir opción de análisis: Datos estadísticos, agrupar y contar
3	Descargar archivo limpio

Tabla 9. Procesos de la Herramienta de Análisis.

datos Pandas. Sus principales características son: su facilidad y el alto rendimiento en el uso de estructuras de datos. Proporciona tres estructuras de datos: Series, Dataframe y Panel. Las primeras, son arreglos unidimensionales con indexación y pueden generarse a partir de diccionarios y listas. Los segundos, son estructuras de datos similares a las tablas de bases de datos relacionales. Los terceros permiten trabajar con más de dos dimensiones (Pandas, 2017).

De acuerdo con el sitio oficial de la librería, sus principales características son: “un objeto DataFrame rápido y eficiente para la manipulación de datos con indexación integrada; herramientas para leer y escribir datos entre estructuras de datos en memoria y diferentes formatos: archivos CSV y texto, Microsoft Excel, bases de datos SQL y el formato HDF5 rápido; alineación inteligente de datos y manejo integrado de datos perdidos: obtenga alineación automática basada en etiquetas en los cálculos y manipule fácilmente datos desordenados en una forma ordenada; recorte inteligente basado en etiquetas, indexación de

grandes conjuntos de datos; está en uso en una amplia variedad de dominios académicos y comerciales, incluyendo Finanzas, Neurociencias, Economía, Estadísticas, Publicidad, Análisis Web y más”.

6.9.6 Herramienta Visualización

Esta herramienta permite subir un archivo xlsx y obtener algunas gráficas básicas para visualizar los patrones encontrarle en tus datos. La tabla siguiente muestra de manera esquemática estos procesos.

Se utilizó para su construcción la librería Matplotlib que es una biblioteca para visualización de datos en segunda dimensión, la cual produce gráficos de calidad para publicar en una variedad de formatos impresos y entornos interactivos a través de distintas plataformas.

Pasos	Visualización de un archivo XLSX
1	Subir archivo
2	Elegir opción de visualización: Barras, histograma y pie
3	Descargar gráfico

Tabla 10. Procesos de la Herramienta de Visualización.

Permite, generar gráficos, histogramas, espectros de potencia, gráficos de barras, gráficos de errores y diagramas de dispersión con solo unas pocas líneas de código (Hunter, Dale, Firing, & Dro, 2017).

Esta librería es versátil y altamente configurable, soporta varios tipos de interfaces gráficas. Se caracteriza por ser simple e intuitiva, de buen rendimiento, de producción profesional y altamente configurable y adaptable a las necesidades específicas de los usuarios (Vaingast, 2014).

Matplotlib permite crear contenido dinámico para visualizarse en un navegador web, embeber gráficos en una Interfaz Gráfica de usuario (GUI) para interactuar con los datos o exportar las salidas generadas a distintos formatos como: JPEG, PNG, PDF y PostScript. Otra manera de utilizar el paquete es por medio de la línea de comandos, opción recomendable para las etapas de desarrollo de aplicaciones (Vaingast, 2014).

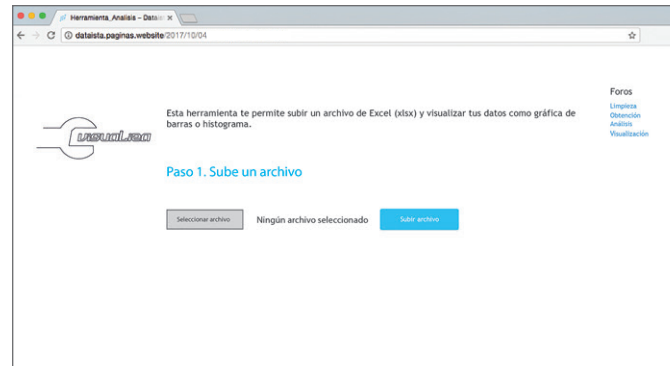


Figura 70. Interfaz de la Herramienta de Visualización.

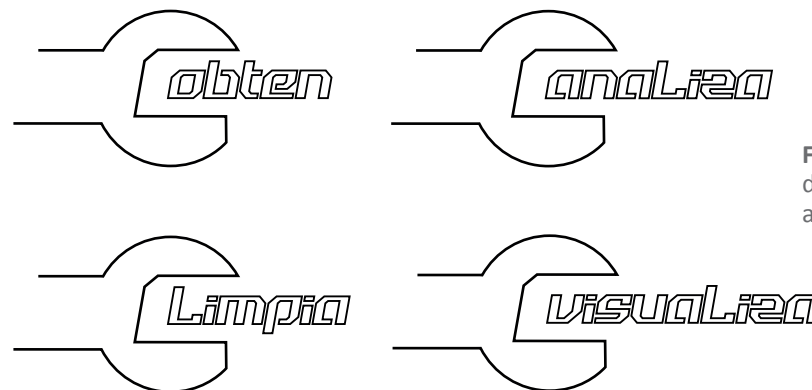


Figura 71. Logos de herramientas auxiliares.

6.10 Resumen del capítulo 6

Al igual que en distintos campos de la actividad humana, en el P.D., la computadora se ha convertido en una herramienta fundamental para: extraer, descubrir, contrastar, combinar, analizar y visualizar datos en información de distintas fuentes. Para ello, es preciso combinar conocimientos y habilidades específicas de distintas disciplinas como lo son: las ciencias de la comunicación, el diseño y la computación, por mencionar algunas. Esta integración busca transformar los datos provenientes de bases de datos en historias a partir de la extracción, filtrado, contextualización, comparación, organización y la jerarquización de datos (Rodríguez Brito & García Chico, 2013).

Para explotar el poder de procesamiento de los sistemas informáticos y las herramientas de cómputo en beneficio del P.D. se necesita tener noción de los siguientes elementos: **1)** los procesos de digitalización, **2)** la filosofía de los datos abiertos, **3)** la transparencia, el acceso público a la información, **4)** nociones estadísticas, **5)** herramientas de visualización, **6)** capacidad para la investigación y **7)** habilidades en el manejo de datos.

En opinión de Flores Vivar (2012) la principal aportación de la computación al P.D. radica en el diseño, programación e implementación de secuencias de comandos para facilitar la realización de tareas complejas en materia de manejo de datos. Ello, con la finalidad de facilitar a los

periodistas y a los usuarios la obtención, limpieza, análisis y visualización de datos y recursos de información. Bradshaw y Rohumaa (2011) señalan la importancia de construir herramientas intuitivas capaces de utilizarse en diferentes plataformas, en el mismo sentido se debe liberar a los periodistas de: **1)** el manejo directo de lenguajes de programación e **2)** inscribirse en costosos y engorrosos cursos de actualización, tanto presenciales como en línea.

Existe en Internet una proliferación de herramientas centradas en la resolución de problemas concretos en el tratamiento de datos, específicamente para recolectar, analizar y visualizar datos. Sin embargo, en la mayoría de los casos la curva de aprendizaje es muy alta para los periodistas que se inician en el P.D. Por ello, la perspectiva de resolución de problemas no es la más adecuada si, la finalidad es alfabetizar a los usuarios en el manejo de datos, pues no se fortalecen las habilidades y conocimientos para enfrentar problemáticas en el uso de datos (Bhargava & D'Ignazio, 2016).

Dataísta nace con la necesidad de construir un sistema de información instruccional para facilitar el entendimiento de las técnicas del P.D.: obtención, limpieza, análisis y visualización. Por esta razón, se desarrolló un prototipo estructurado en torno a los siguientes recursos: **1)** vídeos instruccionales, **2)** un sistema de clasificación de herramientas, **3)** foros, **4)** nube de *tags* o etiquetas, **5)** preguntas frecuentes, **6)** herramientas auxiliares, **7)** listado de bases de datos abiertos y **8)** casos de éxito de P.D. alrededor del mundo.

Todos los apartados anteriores están organizados e interrelacionados dinámicamente para proporcionar a los usuarios acceso a información relevante.

Dataísta se construyó en dos fases: el desarrollo de la plataforma web y la programación de las herramientas auxiliares. En ambos casos se buscó utilizar tecnologías abiertas, es decir, aquellas sin restricciones económicas, legales o tecnológicas para ser utilizadas. La principal ventaja de utilizar esta clase de herramientas es la gran cantidad de proyectos y desarrolladores que pueden ayudar a la continuidad de un proyecto a través de foros, tutoriales, videos y documentos en Internet.

En el desarrollo de la plataforma web, se utilizó WordPress, el cual es un software de gestión de contenido que cuenta con las siguientes características:

- 1) es de código abierto,
- 2) no requiere de un pago,
- 3) es una de las plataformas más utilizada para la publicación en línea,
- 4) existen gran variedad de proyectos que son compartidos por la comunidad haciendo accesible su código,
- 5) tiene la posibilidad de crear diseños personalizados con características tecnológicas robustas,

6) tiene una amplia comunidad de usuarios y desarrolladores que pueden apoyar u orientar durante el proceso.

En el caso de las herramientas auxiliares se utilizó el lenguaje de programación Python y sus principales librerías para el tratamiento de datos como: BeautifulSoup, Pandas, Pygal, Matplotlib y Flask para adaptar los programas en Python al entorno de la web. Python es un lenguaje de programación versátil pues, así como es posible escribir *scripts* de análisis de datos, también se pueden desarrollar aplicaciones web. Entre sus principales características están:

- 1) su extensibilidad,
- 2) ser software libre,
- 3) su compactibilidad y
- 4) ser un estándar.

Una de las grandes aportaciones de Dataísta es el uso de videotutoriales mediante los cuales, de una manera fácil, se explica cada una de las etapas del P.D. Así, el usuario será capaz de seguir paso a paso las etapas explorando una herramienta, especialmente, desarrollada para el proceso así como utilizar otras más comerciales.

En el siguiente capítulo se presenta la evaluación de Dataísta, fase que permitió identificar problemas de técnicos y de usabilidad.

Evaluación



Capítulo

Una vez que el sistema estuvo disponible en línea y funcional, el siguiente paso fue evaluar nuestro sistema con los potenciales usuarios de Dataísta.

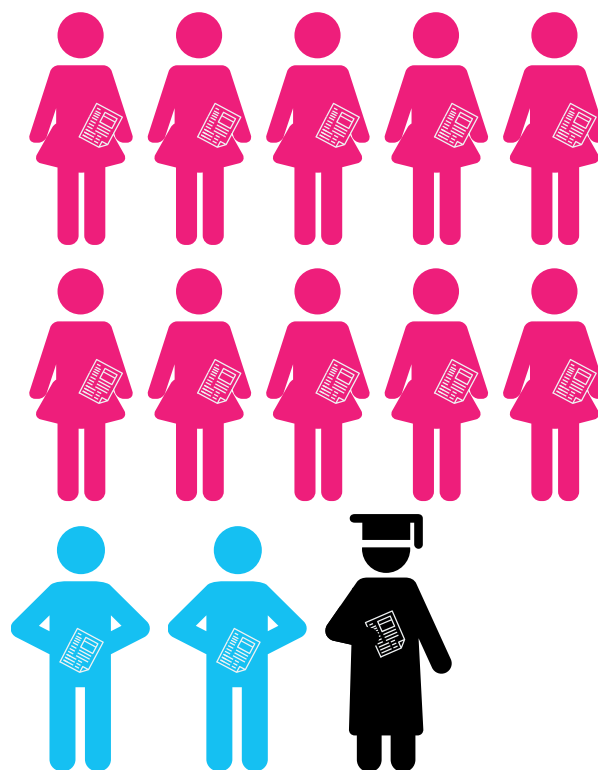
7.1 Pruebas de Usabilidad

Como primera fase, se llevaron a cabo unas pruebas de usabilidad durante el Seminario de Periodismo de Investigación organizado por la “Catedra Miguel Ángel Granados Chapa”, el día 7 de julio de 2017.

Durante estas pruebas, se presentó por primera vez al público el Sistema de Información Dataísta y se invitó a los asistentes a navegar en el sistema con la finalidad de que compartieran sus impresiones.

En esta primera fase, lo principal fue evaluar la navegación, así como su funcionalidad. A continuación, se presenta un listado de los puntos más relevantes mencionados durante la prueba:

13



Personas entre 26 y 30 años, en su mayoría mujeres, periodistas, a excepción de 1 estudiante.

Figura 72. Resultados de la prueba de usabilidad. (Fuente: Elaboración propia)

Hallazgos:



Las herramientas planteadas en cada sección presentaron mayores problemas, principalmente porque carecen de instrucciones de uso.

No queda claro que el logotipo de Dataísta es también una forma de volver a la página de inicio.

No parecen claros los botones del menú principal, por lo que habría que agregar una instrucción.



Los usuarios preferirían que la actualización del catálogo de sitios fuera inmediata, por lo que una explicación de cómo funciona la publicación podría ser buena opción.

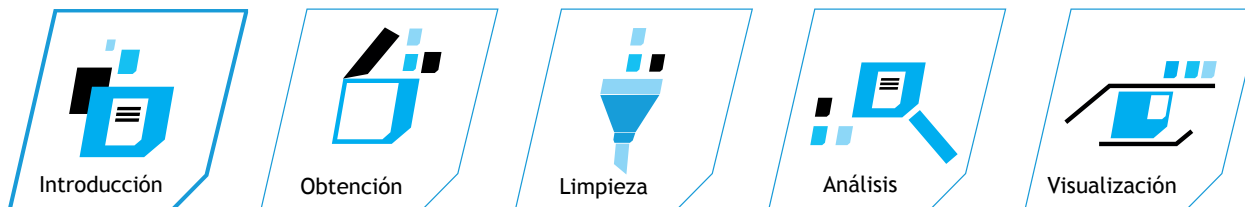


Figura 73. Resultados de la prueba de usabilidad.
(Fuente: Elaboración propia)



Una vez que se recopilaron y agruparon los comentarios de los usuarios, se trabajaron los ajustes pertinentes para poder someter nuevamente al sistema a evaluación, esta vez no solo considerando los aspectos de usabilidad sino en general todo el funcionamiento de Dataísta.

7.2 Sesión de *Focus Group*

La segunda fase de la evaluación consistió en la realización de un *focus group* con usuarios potenciales del sistema (periodistas de investigación). Para ello se lanzó una convocatoria a través de redes sociodigitales (Facebook y Twitter) y correos electrónicos a los diferentes periodistas que participaron con nosotros en las etapas previas de la investigación.

La sesión se llevó a cabo el día 9 de septiembre de 2017, en “Casa Rafael Galván” de la UAM con una asistencia de 10 personas. La agenda estuvo distribuida de la siguiente manera:

- 10:00 - 10:15 Recepción
- 10:15 - 11:30 Bienvenida / Presentación General de Dataísta
- 11:30 - 12:00 Interacción usuarios-Dataísta

- 12:00 - 13:00 Discusión
- 14:00 Conclusiones y cierre del evento

Como se menciona en la agenda, se invitó a los usuarios a probar el sistema, tuvieron media hora para interactuar con ella sin ninguna intervención. Una vez que pasó el tiempo se procedió a la discusión siguiendo los siguientes puntos:

- Usabilidad
 - Navegación
 - Claridad en el *journey* (recorrido)
 - Botones
 - Facilidad de uso
 - Compatibilidad con dispositivos
- Contenidos y Utilidad
 - Claridad en las explicaciones de los videos
 - Orden de ideas
 - Vínculo entre fases
 - Vínculo entre audio y video

■ ¿Información suficiente? ¿Podrían empezar a analizar Datos Abiertos con los temas revisados?

■ ¿Es como lo imaginaban?

■ Lógica de la metodología

■ Sugerencias

Como resultado de la discusión se procedió a analizar la información, misma que clasificamos bajo los siguientes rubros:

Generales

En esta categoría, se incluyeron aquellos comentarios que los integrantes del equipo fueron escuchando y anotando durante el tiempo que los usuarios fueron interactuando con el sistema.

■ No queda muy claro cuál es el *home*, si la pantalla inicial donde se tiene el video explicativo o bien, la pantalla donde está el menú principal de la metodología.

■ El *login* es confuso. Al registrarte, da la opción de “iniciar sesión” y viceversa.

■ Es necesario revisar la compatibilidad entre dispositivos y navegadores.

■ La diferencia entre menús no es clara, ¿valdría la pena hacer una animación que

explique cómo usar el sitio?

■ Las herramientas auxiliares para cada fase de la metodología carecen de instrucciones claras.

■ Valdría la pena incluir una subclasificación de herramientas, es decir, no solo por la dificultad de uso, sino por su objetivo. P.e: búsqueda, *scraping*, visualización por mapas, etc.

■ La parte de herramientas necesita de información un poco más detallada, decir “requiere de conocimientos en programación es cruel...” El usuario sale de la herramienta explorada al percatarse que los conocimientos que necesita son complejos. Se recomienda una clasificación por *tags*.

■ Dejar claro que las herramientas sugeridas y las fases de la metodología son diferentes.

■ El video explicativo solo pasa una vez. Si el usuario quisiera consultarlo de nuevo, no tiene dónde buscarlo.

■ Al haber separado la sección de “Datos Abiertos” del resto del sistema, le falta una introducción que explique qué son.

■ En el formulario para agregar bases de datos abiertas al sistema, se solicita el

nombre. Hay que especificar si es el nombre de la base de datos o del usuario, no es claro. Los usuarios no encuentran sentido a la acción de “pedirles su nombre” nuevamente para sugerir bases de datos.

■ Se recomienda inducir a la interdisciplina desde el inicio.

■ Nunca se habla de los creadores o se proporciona algún contacto.

■ Se sugiere ligar otros cursos y obtener una comisión por las plataformas referidas.

■ El tamaño de la letra parece pequeño, la tendencia es grande, 16px o al menos 14 px.

■ Se sugiere hacer un plan de difusión de Dataísta.

Usabilidad

A partir de este rubro se mencionan los comentarios derivados de la discusión después de que los usuarios exploraron el sistema.

■ Les parece bueno que el contenido se centre en la metodología del P.D., sin embargo, no queda clara la relación del menú superior con esta. De conservarse el menú superior, la sugerencia es agregar algún texto introductorio. La recomendación general es simplificar

el diseño con un mínimo de botones y enfocarse solo en la metodología.

■ El menú vertical desaparece en la versión móvil.

■ Una vez que el usuario se registró, el botón de “iniciar sesión” sigue apareciendo una vez iniciada la sesión. No cambia por el nombre del usuario ya registrado.

■ Una vez que se hace el registro, de alguna manera debería indicar que el usuario está dentro del sistema.

■ No es clara la utilidad/aportación del menú superior. Además de que, contiene demasiada información.

■ El CTA (call to action) del video de obtención, envía al usuario a dar clic en el botón debajo del video. Sin embargo, en el sitio el botón se fue hasta abajo (no todos lo encontraron). Se sugiere jerarquizar el botón “Ir a scraping” de otra manera.

■ El botón de home está ubicado de forma rara (esquina superior derecha, lo usual es que esté localizado del lado izquierdo).

■ Simplificar la cantidad de pasos. Muchas veces en el menú izquierdo se requieren de muchos clics para llegar al contenido deseado.

- Revisar la compatibilidad en diferentes navegadores, no en todos los dispositivos carga de manera uniforme.
- En algunos navegadores la sección de “herramienta auxiliar” no carga de manera apropiada.
- Vale la pena, replantear la estructura de los menús. El menú izquierdo es difícil de digerir, mientras que, es mucho más natural el menú horizontal.
- Al usar el *scroll* y navegar por la parte inferior, puedes fácilmente perder la ruta.
- Se sugiere abrir enlaces en una ventana nueva, ya que es incómodo abrirlos en la misma.
- Los foros no pueden verse en los celulares o tablets.

Contenido

- Sería bueno, incluir una infografía para consulta posterior por cada fase de la metodología, siguiendo una estructura como: Texto – Video – Infografía descargable.
- En la explicación de las herramientas, sería útil incluir una subclasificación que permita identificar no solo el nivel de facilidad de uso (como usuario), sino el propósito de la herramienta.
- La explicación de las herramientas, en algunos casos es ambigua. P.e, en una dice que sirve para convertir archivos PDF a otros formatos.
- ¿Por qué en la sección de medios se separa a LATAM de US/Canadá? En esta sección no es clara la división por continentes.
- Explicar cómo se decidió tomar como punto de partida esa metodología (incluir las fuentes consultadas). Esto podría ser después de explicar qué es Dataísta (Este sitio es resultado de...) y si hay una autoridad colegiada, habría que incluirla. Si bien se incluye en las preguntas frecuentes, es necesario darle mayor peso.
- Sería ideal incluir una sección después de visualización donde se aborde la parte de narrativas. Sería una buena forma de dar cierre a todo el proceso. Se recomienda explorar el sitio narrativescience.com
- En la sección de análisis, profundizar en la parte de correlación/causalidad. Se sugiere presentar más opciones en los videos o infografías con este ejemplo,

causa – consecuencia u otros razonamientos.

■ Comunicar que esta metodología de trabajo, también puede ser funcional para la labor del día a día del periodista (no siempre tienen que ser investigaciones complejas/robustas), acercar las técnicas a los periodistas. Los medios requieren la información procesada y es costoso, se podría pensar en incluir la herramienta a algún periódico o en que cierta redacción ingresara a Dataísta, realizara un trabajo y lo publicara, facilitando el trabajo a los medios de comunicación, una manera a su vez, de monetizar la plataforma.

■ Los foros son un punto de atención, pero podrían ser más visuales, más contemporáneos. Rediseño del foro, luce muy “reto”/”90s”. Por otro lado, Los foros tienen valor, pero aparentemente la línea del proyecto se pierde.

■ Habría que dar herramientas o promover el trabajo colectivo.

La palabra “Medios” no parece pertinente. El usuario piensa más bien en nuestro contacto o redes (sección de la que carece el sistema). Se sugiere pensar en otra denominación, por ejemplo: “mejores prácticas”.

■ La hipótesis no la formulan los periodistas, solo hacen conjeturas. Se siente un poco forzado para los periodistas, quizá sería recomendable modificarlo.

■ Hay cosas que se dan por obvias, por ejemplo, la verificación de la información en la sección “Limpieza”, podríamos integrar la verificación de la información como consejo a la limpieza.

■ Se sugiere trasladar las categorías “Medios” y “Datos abiertos” a la sección de “Obtención”.

Una vez que se enlistaron todos los puntos mencionados por los participantes, se procedió al análisis de la información, para ello agrupamos los conceptos mismos que visualizamos en las siguientes nubes de palabras:



Figura 74. Nube de palabras relacionadas con usabilidad del sistema. (Fuente: Elaboración propia)



Figura 75. Nube de palabras relacionadas con el contenido del sistema. (Fuente: Elaboración propia)

Finalmente, identificamos atributos positivos del sistema dentro de los comentarios que realizaron los usuarios, se considera relevante mostrarlos, pues en esta aproximación a ellos, es como se puede constatar que la hipótesis y los objetivos planteados al inicio del proyecto se ven reflejados en Dataísta.

Explicaciones
Útil breves y concisas
 Íconos hacen ilustrativos a videos **Identidad gráfica atractiva**
 Fomenta el trabajo colaborativo **De largo futuro**
Interesante **Sí utilizarían Dataísta**
 Contenidos centrados en Metodología

Figura 76. Aspectos positivos del sistema. (Fuente: Elaboración propia)

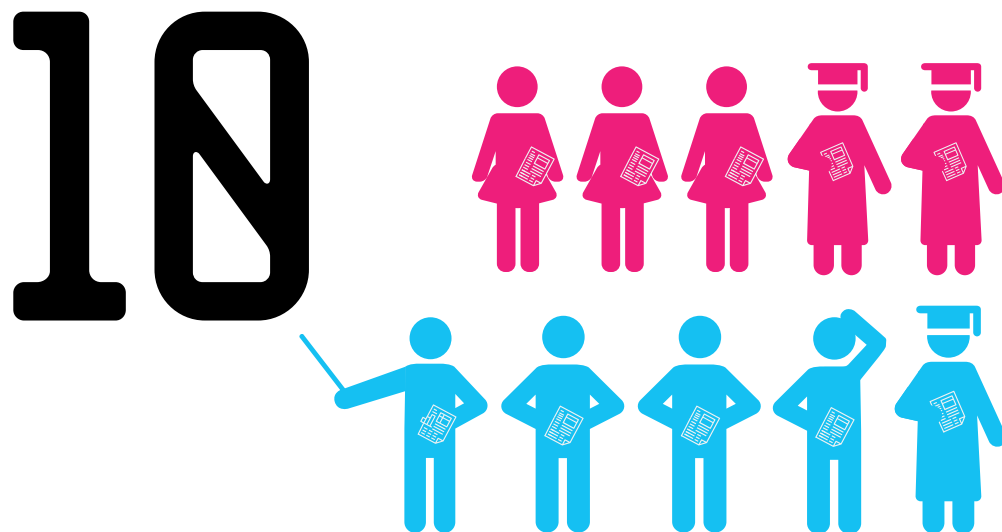


Figura 77. Focus Group “Dataísta”. (Fuente: Elaboración propia)

Mujeres y Hombres, principalmente estudiantes de comunicación y periodistas.

Principales Hallazgos

- de usabilidad: No es clara la jerarquía entre el menú general y el superior. Sería mejor desplegar la herramienta auxiliar en otra página. El desplazamiento con el scroll es muy extenso. No se puede regresar al video introductorio.
- en el contenido: Adicional a los videos, sería útil incluir una infografía para consulta posterior.
- positivos: Los usuarios sí utilizarían Dataísta, lo consideran útil, interesante y de largo futuro, con explicaciones breves y concisas y de identidad gráfica atractiva.

Después de analizar los resultados obtenidos durante el Focus Group, se seleccionaron los siguientes puntos a mejorar en el sitio:

- el problema de reconocimiento de los usuarios registrados quedó resuelto,
- se redujo la cantidad de botones en el menu superior,
- se redujo el scroll en cada una de las secciones referentes a la metodología,
- el botón de “home” cambió de lugar a la esquina superior derecha, de forma en que los usuarios puedan ubicarlo rápidamente,
- en la sección de “Preguntas Frecuentes” se integraron los siguientes elementos;
- información de contacto,
- acceso al video introductorio,
- explicación del nombre del Sistema,
- en la clasificación de las herramientas, los links se abren en nuevas pestañas del navegador,
- la sección de medios fue replanteada, de tal forma que su exploración sea mucho más amigable con el usuario.

La selección de los ajustes listados anteriormente se realizó con base en la relevancia que tienen estos cambios en la experiencia del usuario de “Dataísta”, pues el principal objetivo es contribuir en su camino al aprendizaje de la metodología del P.D., por lo que si el Sistema no es intuitivo o fácil de usar, es probable que no vuelvan a consultarlo. Con los ajustes implementados, se ha puesto en línea el producto final, disponible en: www.dataista.org.

7.3 Resumen del capítulo 7

Luego de presentar Dataísta ante 13 personas entre 26 y 30 años, en su mayoría mujeres y periodistas, a excepción de un estudiante, asistentes del Seminario de Periodismo de Investigación organizado por la “Catedra Miguel Ángel Granados Chapa”, con el objetivo de evaluar su usabilidad en una primera fase, destacó la necesidad de proporcionar mayores instrucciones, principalmente en el uso de las herramientas.

Una vez realizados los ajustes y después de lanzar una convocatoria dirigida a los diferentes periodistas que participaron previamente en la investigación, el sistema fue sometido a una segunda fase de evaluación en un *focus group* constituido por diez usuarios potenciales del sistema: estudiantes de comunicación y periodistas.

Se presentó el sistema y se invitó a los usuarios

a interactuar con él sin intervención de los maestros, se observaban y registraban los detalles relevantes para posteriormente discutir no solo la usabilidad, sino también los contenidos y todo aquello que constituye Dataísta.

La información obtenida como resultado de la discusión, fue clasificada bajo los rubros: Generales, Usabilidad y Contenido. En el primero, incluimos las observaciones registradas durante la interacción de los usuarios con el sistema. Mientras que, en el segundo y tercer rubros, los comentarios derivados de la discusión.

Una vez enlistadas las sugerencias de los participantes, sintetizamos la información, agrupando los conceptos en nubes de palabras y la analizamos con la intención de implementar en el sistema aquello que resultó de mayor relevancia para el usuario. En el caso de la Usabilidad, destaca la relación jerárquica entre el menú general y el superior, la pertinencia de una página para la herramienta auxiliar, el extenso desplazamiento del *scroll* y el libre acceso al video introductorio. Mientras que, para el contenido, destaca la inclusión de infografías complementarias a los videos, la reconsideración de los foros y en general, profundizar más en la metodología y herramientas.

Se decidió mostrar, además, los atributos positivos del sistema identificados en los comentarios, pues conforman también aspectos cualitativos de la evaluación.

A continuación, se presentan las conclusiones a las que llegó el equipo, una vez que se concluyó el proyecto.

Conclusiones

Capítulo **8**

Si bien es cierto que el Periodismo siempre se ha caracterizado por su función social y ha utilizado datos en sus investigaciones, el entorno digital parece estar logrando una transición social mayor hacia el P.D. Las grandes cantidades de información y, específicamente los Datos Abiertos, son una promesa latente de grandes cambios sociales.

El P.D. es la forma en la que el periodismo se adapta a las condiciones digitales y aprovecha tanto los grandes repositorios de información digital, así como la necesidad de las sociedades por solicitar mayor transparencia a los gobiernos y el potencial ejercicio de derechos ciudadanos.

La información que se transmite a través de los medios de comunicación, con el formidable apoyo de las nuevas tecnologías, tiene una importancia decisiva para el desarrollo individual y social. Es indispensable para la vida democrática, ya que la participación de los ciudadanos en los asuntos públicos es de vital importancia para que pueda desarrollarse plenamente la democracia en, por ello debe estar garantizada. Esta participación sería imposible si los ciudadanos no reciben la información pública necesaria que deben procurarle los medios de comunicación (Villanueva, 2004, pág. 175).

Actualmente, los periodistas de investigación no cuentan con los conocimientos necesarios para hacer uso de los Datos Abiertos que tanto el gobierno mexicano, como otras instituciones ponen a disposición de los usuarios en línea. Aunque podrían ser autodidactas, consultar libros y

acceder a las herramientas disponibles en la red, la gran mayoría de las veces no pueden llegar al entendimiento de la metodología del P.D., ya que desconocen los conceptos básicos relacionados con otras disciplinas que también la nutren, como lo son el Diseño y la Computación.

Además, la falta de tiempo y el nulo apoyo de las redacciones donde laboran, impiden que aquellos que están interesados en esta especialización puedan hacerlo. Por ello, surge la necesidad de un sistema de información, en este caso Dataísta, que esté destinado y pensado en las necesidades específicas de los periodistas mexicanos. Esto, permitirá reducir la barrera relacionada con la formación académica, hasta que las instituciones educativas y los medios de comunicación se den cuenta de la importancia que tienen en invertir en esta nueva forma de hacer periodismo y el P.D. forme parte de los currículos.

La finalidad de Dataísta es combatir las problemáticas arriba mencionadas para permitir a los periodistas de investigación entender y utilizar el razonamiento propio de la metodología desarrollada en el P.D. para articular sus investigaciones a través de las diferentes bases de D.A. El Sistema de Información combina las fortalezas de educación en línea y de un curso masivo en línea (MOOC). De los primeros retoma, el uso de un entorno cerrado de aprendizaje, la comunicación a través de foros de discusión y el registro previo en el sistema para utilizar algunas características. De los segundos, el acceso gratuito al contenido, la finalidad de apoyar a una comunidad y el énfasis

en el proceso de aprendizaje más que en la evaluación.

Para su diseño y construcción como Sistema de Información, se desarrolló un prototipo estructurado en torno a los siguientes recursos:

- 1) vídeos instruccionales,
- 2) un sistema de clasificación de herramientas,
- 3) foros,
- 4) nube de tags o etiquetas,
- 5) preguntas frecuentes,
- 6) herramientas auxiliares,
- 7) listado de bases de D.A. y,
- 8) casos de éxito de P.D. alrededor del mundo. Todos los anteriores están organizados e interrelacionados dinámicamente para proporcionar a los usuarios acceso a información relevante.

Dataísta se desarrolló en dos fases: la creación de la plataforma web y la programación de las herramientas auxiliares. En ambos casos se buscó utilizar programas abiertos, es decir, aquellos sin restricciones económicas, legales o tecnológicas para ser utilizados. La principal

ventaja de utilizar esta clase de herramientas es la gran cantidad de proyectos y desarrolladores que pueden ayudar al desarrollo de un proyecto a través de foros, tutoriales, videos y documentos en internet.

Como se ha venido mencionando en todo el documento, la base que sustenta el Sistema de Información Dataísta, es la metodología propuesta por el P.D., tomándolo como base se pudieron definir las diferentes secciones que conforman el Sistema. A partir de ello y el cruce de la información recopilada durante la observación participativa, se establecieron las demás secciones que lo conforman.

Dataísta se caracteriza, además de mostrar la metodología del P.D, por el uso de videos infográficos que permiten comunicar y generar un proceso de enseñanza-aprendizaje no solo auditivo sino también visual. Los videos contienen características importantes para poder lograr la comprensión y aprendizaje de los contenidos. Se diseñaron a partir de un guion que permitió:

- Clarificar ideas.
- Diseñar con un propósito.
- Sintetizar en pocos minutos el tema.
- Utilizar un lenguaje claro y conciso.

Así, se considera que una vez que el periodista de investigación pone en marcha el video es

capaz de entender en pocos minutos cada una de las etapas, que forman la metodología del P.D., de una forma lúdica.

Es importante destacar que, dentro de los contenidos del Sistema, se hizo especial énfasis en la importancia de trabajar de manera conjunta con Diseñadores y Programadores; esto porque las investigaciones basadas en datos requieren de la exploración y el trabajo colaborativo de disciplinas como Comunicación, Diseño y Computación.

Dataísta, ayudar al periodista a tener un lenguaje común con otras las otras disciplinas, de tal forma que se puedan desarrollar investigaciones robustas que permitan aprovechar la información disponible. No se pretende que el periodista se convierta en un diseñador o programador experto, sino que comprenda los procesos y razonamientos que hay detrás de la metodología del P.D., para que así pueda colaborar con otras disciplinas.

Otro punto a destacar del Sistema, es la incorporación de un catálogo de herramientas auxiliares en cada una de las fases de la metodología; cuenta con una clasificación por niveles de conocimiento, lo cual permite que el usuario identifique el nivel de conocimientos que posee y con base en ello seleccione la que más se adapte a sus necesidades y competencias.

Además, se incluyó en cada una de las fases que conforman la metodología el P.D., una herramienta auxiliar basada en el paradigma de los sistemas concisos; la idea de su inclusión es que el periodista inexperto pueda empezar a explorar sus bases

de datos a partir del uso de herramientas sencillas que le permitan comprender y vincular los procesos que deben llevarse a cabo a la hora de realizar una investigación basada en datos.

Existe en internet una vasta oferta de herramientas para el tratamiento de datos, desde robustas soluciones de pago, pasando por experimentos académicos, hasta soluciones de software libre. La mayor parte de ellas hace énfasis en la resolución de problemas concretos sin dar opciones al usuario para entender el proceso de trabajo.

De manera precisa las herramientas auxiliares permiten al periodista iniciarse en tareas básicas en el tratamiento de datos. Por ejemplo, la obtención de tablas de archivos PDF, eliminar caracteres, palabras y transformar mayúsculas y minúsculas y viceversa en archivos XLSX. También es posible analizar archivos XLSX desde un punto de vista estadístico, agrupar y contar datos para conseguir un panorama general de un archivo. Así como crear visualizaciones básicas a partir de archivos XLSX, ya sean gráficas de barras, histogramas y de pie.

Durante la investigación se identificó que la comunidad *datera* en México es muy activa no solo compartiendo su trabajo, sino trabajando de manera colaborativa frecuentemente. Organizaciones como *Escuela de Datos*, *SocialTIC* y *Chicas Poderosas* han contribuido a que periodistas, diseñadores, programadores y científicos de datos trabajen en conjunto en diferentes proyectos, compartan su trabajo o experiencias. Por ello, también se decidió incluir un foro dentro

del Sistema, su finalidad es que los usuarios de Dataísta se conozcan, intercambien experiencias que nutran sus procesos de aprendizaje o incluso trabajen de manera colaborativa. El diálogo con los otros, es una característica de los trabajos interdisciplinarios y, como se mencionaba en párrafos anteriores, es vital para el óptimo desarrollo de este tipo de investigaciones.

La innovación de Dataísta reside en que es el único sistema, hasta la fecha, que introduce al periodista de investigación en el P.D. a través del acercamiento a otras disciplinas. En este caso, el uso de herramientas informáticas son parte esencial de la propuesta de Dataísta. Asimismo, aporta un proceso de enseñanza-aprendizaje para usuarios con diferentes conocimientos y habilidades logrando la introducción de estos y la puesta en marcha de trabajos especializados en el P.D.

Dataísta es un sistema basado en el Diseño Instruccional incorporando a sus usuarios desde la concepción hasta su implementación, detectando necesidades claras del periodista de investigación en su aprendizaje sobre la metodología del P.D.

Al estar enmarcado dentro de un proceso de Diseño Centrado en el Usuario, Dataísta ha sido evaluado y seguido de cerca por periodistas interesados en utilizarlo de manera activa. Por el momento, la evaluación de Dataísta ha permitido mejorar aspectos de usabilidad que ayudarán a que el sistema sea más fácil de usar por el usuario no experto en aspectos informáticos, relacionados de manera inherente con la percepción de los usuarios al reconocer las dificultades

en su proceso de interacción con el sistema. Como resultado, Dataísta cuenta con las siguientes instrucciones que permiten:

- Identificar el menú principal para comenzar a explorar las etapas utilizadas en la metodología del P.D. o regresar a la Introducción.
- Entender el propósito de cada una de las herramientas.
- Regresar fácilmente al inicio.
- Además, los botones se han reducido al mínimo para conservar la relación de la jerarquía del menú principal con las etapas de la metodología.
- El despliegue de las herramientas auxiliares en páginas propias, ha reducido el tamaño de las mismas y el desplazamiento del usuario en ellas, con el *scroll*.

Adicionalmente, como parte del contenido y complemento de los videos, se trabaja también en infografías descargables para consulta del usuario.

La evaluación del sistema durante los diferentes momentos de su desarrollo, permitió que el producto final disponible en www.dataista.org, sea funcional y útil para los periodistas de investigación que lo consulten.

La implementación de Diseño Instruccional y Diseño Centrado en el Usuario, permitió que

Dataísta se adecua a las necesidades de los periodistas de investigación mexicanos que quieren incursionar en el P.D., pues no solo se sustenta en información documental y experiencias de aquellos periodistas mexicanos que ya hacen investigaciones aplicando la metodología del P.D., sino que también contempla las necesidades de los usuarios finales, haciéndolo accesible y sobre todo útil para aquellos interesados en incursionar en el P.D.

Durante la investigación, se confirmó la importancia que el P.D. y los D.A. han adquirido como grandes emisores de la sociedad en el entorno digital.

Creemos que la creación de Dataísta a partir de la identificación de las dificultades que enfrenta el país, el periodista de investigación en su incursión al periodismo de datos; desde su formación a la integración de equipos interdisciplinarios, si bien está distante de ser una solución de problemáticas que salen de nuestro alcance, es una manera de contribuir para explotar las oportunidades que esta especialización ofrece e incrementar la cantidad de casos que puedan generar interactividad de la información con la sociedad y a su vez, conducir hacia cambios mayores.

Si el gobierno mexicano, realmente quiere obtener provecho de las ventajas tecnológicas para impulsar mayores cambios sociales, además de poner los Datos Abiertos al alcance de la población, deberá tomar medidas en, la renovación de la formación profesional (no

únicamente la de los periodistas), el financiamiento para equipos interdisciplinarios que puedan convertirse en gestores de conocimiento, la eliminación de obstáculos al proporcionar la información, y la censura política.

Dataísta fue elaborado dentro de un equipo interdisciplinario mostrando, también, que el P.D. necesita fuertemente de la colaboración de distintas disciplinas. En este caso, el Diseño, las Ciencias de la Comunicación y la Computación originaron a Dataísta que es un sistema para el P.D. que tiene que trabajar de manera interdisciplinaria. Con esto, no pensamos que el P.D. deba de cambiar de disciplina, sino que debe de entender los procesos que llevan a cabo cada una para poder trabajar conjuntamente y obtener resultados pertinentes que no se darían solo con una disciplina.

Por otra parte, Dataísta seguirá siendo evaluado, monitoreado y mejorado con el apoyo de la comunidad de periodistas de datos que lo utilizarán en su proceso para comprender la metodología del P.D.

Referencias

- Universidad Carlos III de Madrid. (17 de Mayo de 2017). Guía del profesor para la planificación, diseño e impartición de MOOCs. Obtenido de http://docubib.uc3m.es/MOOCs/Guia-metodologica-MOOC-Wimba/page_06.htm
- Ackerman J, C. (2008). Más allá del acceso a la información, transparencia, rendición de cuentas y estado de derecho. Ciudad México, México: Siglo XXI editores S.A de C.V en coedición con, Instituto de Investigaciones Jurídicas de la UNAM Universidad de Guadalajara, Cámara de Diputados, Centro Internacional de Estudios sobre Transparencia y la Información.
- Aggarwal, S. (2014). *Flask Framework Cookbook*. Birmingham: Packt Publishing.
- American Press Institute. (s.f.). Recuperado el 17 de 01 de 2017, de *Journalism Essentials*: <https://www.americanpressinstitute.org/journalism-essentials>
- Análisis y Diseño Orientado a Objetos. (02 de 04 de 2017). Obtenido de Osiris LMS.
- Animal Político. (23 de 10 de 2015). *NarcoData*. Recuperado el 24 de 10 de 2017, de <http://narcodata.animalpolitico.com>
- Arruti, A. (2001). *Ciberperiodismo. Nuevos enfoques, conceptos y profesiones emergentes en el mundo infodigital*. Madrid, España: Editorial Limusa, S.A de C.V. Grupo Noriega Editores.
- Artículo 19. (6 de 10 de 2017). Recuperado el 25 de 09 de 2017, de <https://articulo19.org/periodistasasesinados>
- Assange , J., & Romero, P. (2013). *Criptopunks. La libertad y el futuro de internet*. *Revista Austral de Ciencias Sociales*, 24, 151-156.
- Ávila, A. M. (2013). *Periodismo de Datos: Historia y momento actual*. En F. Perry, & M. Paz, *Manual de Periodismo de Datos Iberoamericano* (pág. 6).
- Barbour, R. (2013). *Los grupos de discusión en Investigación Cualitativa*. Madrid: SAGE Publications.
- Barranco, R. F. (2012). *¿ Qué es big data? Information Management*.
- Béjar, E. (22 de enero de 2012). *5 * de los Datos Abiertos*. Recuperado el 1 de diciembre de 2015, de *5 * de los Datos Abiertos*: <http://5stardata.info/es/>
- Bhargava, R., & D'Ignazio, C. (2016). *Designing Tools and Activities for Data Literacy Learners*.
- Boyd, D., & Crawford, K. (2012). *Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon*. *Information, communication & society*, 15(5), 662-679.
- Bradshaw, P., & Rohumaa, L. (2011). *The Online Journalism Hadbook: Skills to Survive and Thrieve in the Digital Age*. Londres: Longman Practical Journalism Series.
- Bradshaw, P., & Rohumaa, L. (2011). *The Online Journalism Hadbook: Skills to Survive and Thrive in the Digital Age*. Londres: Longman Practical Journalism Series.
- Bravo, A. A. (2017). *El Periodismo de Datos y la Web Semántica*. *Cuadernos de Información y Comunicación (CIC)*, 18, 99-116.

- C. B., & C. P. (2016). *Teaching, Data and Computational Journalism*. Nueva York, NY: Columbia Journalism School. Recuperado el 13 de Mayo de 2017, de https://journalism.columbia.edu/system/files/content/teaching_data_and_computational_journalism.pdf
- Calderón, F. (14 de 05 de 2014). Forbes México. Recuperado el 05 de 03 de 2017, de <http://www.forbes.com.mx/las-5-tendencias-del-aprendizaje-en-linea/#gs.itxPqCk>
- Carvajal, R. (2016). Qué es el scraping y cómo hacerlo bien. En F. Perry, & M. Paz, *Manual de Periodismo de Datos Iberoamericano*. PoderOmedia.
- Casanueva, J. (Octubre de 2015). El periodismo y la agenda de Datos Abiertos en América Latina. Obtenido de Red de Periodistas Internacionales: <https://ijnet.org/es/blog/el-periodismo-y-la-agenda-de-datos-abiertos-en-am%C3%A9rica-latina>
- Cazaux, D. (2016). *Sociedad Civil, Ética y Periodismo Científico. Razón y Palabra*. Recuperado el 21 de 11 de 2016, de http://www.razonypalabra.org.mx/periodismo_cientifico/dcasaux3.html
- Contreras, P. (2003). *Me llamo Kohfam. Identidad hacker, una aproximación antropológica*. Gedisa.
- Córdova, D. (2002). *El Diseño Instruccional: Dos Tendencias y una Tansición Esperada*. Docencia Universitaria, III.
- Creswell, J. (2014). *Research Design: qualitative, quantitative and mixed methods approaches*. Thousand Oaks: SAGE Publications.
- Crucianelli, S. (2013). ¿Qué es el Periodismo de Datos? Cuadernos para Periodistas, 106-124.
- Crucianelli, S. (2013). *Herramientas Digitales para Periodistas*. Texas: Knight Center for Journalism in Ameerica
- Dallal, A. (1976). Periodismo y Literatura. *Diálogos: Artes, Letras, Ciencias Humanas*, 29-31.
- Diario Oficial de la Federación. (s.f.). Decreto por e que se establece la regulación en materia de Datos Abiertos. Obtenido de http://www.dof.gob.mx/nota_detalle.php?codigo=5382838&fecha=20/02/2015
- Díaz, N. F. (20 de 06 de 2017). La Libertad de Expresión en México. Recuperado el 10 de 09 de 2017, de <https://lasillarota.com/opinion/columnas/la-libertad-de-expresion-en-mexico/158356>
- Dorrego, E. (2016). Educación a distancia y evaluación del aprendizaje. *Revista de Educación a Distancia*. Obtenido de <http://revistas.um.es/red/article/view/271241/198481>
- Dwyer, G. (2016). *Flask by Example*. Birmingham: Pack Publishing.
- Dwyer, G. (13 de Febrero de 2017). Flask vs. Django: Why Flask Might Be Better. Recuperado el 10 de Septiembre de 2017, de Codementor Community: <https://www.codementor.io/garethdwyer/flask-vs-django-why-flask-might-be-better-4xs7mdf8v>
- El País. (28 de Enero de 2014). Obtenido de http://internacional.elpais.com/internacional/2014/01/28/actualidad/1390941981_522546.html
- El Universal Data. (2015). *Desaparecidos*. Recuperado el 24 de 10 de 2017, de <http://interactivo.eluniversal.com.mx/desaparecidos/index.html>
- Elías, C. (2015). ¿Wikileaks es periodismo ciudadano? De la ética del hacker del Cablegate a la estética emo de Assange como ícono global. *Revista Latina de Comunicación Social*.

- Eriksson, H. E. (2004). UML Toolkit. Wiley Computer Publishing.
- Fabila, R. A. (2007). La Censura al Periodismo en México: Revisión Histórica y Perspectivas. *Razón y Palabra*, 6.
- Félix Anduaga, J. A., Alvarado Noriega, E., & Mendivil Alvarado, L. (2015). Censura y autocensura en la cobertura de temas relacionados con el narcotráfico y la delincuencia organizada en el Estado de Sonora, México. *Communication Journal*, 129-160. Recuperado el 7 de 08 de 2017
- Flick, U. (2009). *An Introduction to Qualitative Research*. Thousand Oaks, California: SAGE Publications.
- Flores Vivar, J. (2012). Ecosistema del Periodismo de Datos: como combinar fuentes, datos, contenidos, tecnología y mashups en el desarrollo del data journalism en la red. *Comunicacao y Sociedade*, 34, 7-35.
- Flores, J. V., & Salinas, C. A. (2013). El periodismo de datos como especialización de las organizaciones de noticias en Internet. *Correspondencias y Análisis*, 15-34.
- Forbes. (2015). Forbes. Obtenido de <http://www.forbes.com.mx/mexico-el-pais-mas-peligroso-de-america-latina-para-periodistas/>
- Freedom House. (s.f.). Recuperado el 20 de 09 de 2017, de <https://freedomhouse.org/report/freedom-press/2016/mexico>
- Gantz, J., & Reinsel, D. (2012). The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east. IDC iView: IDC Analyze the future, 1-16.
- Garduño, R. V. (2005). *Enseñanza Virtual: Sobre la Organización de Recursos Informativos Digitales*. México: UNAM: Centro Universitario de Investigaciones Bibliotecológicas y de la Información.
- Global Investigative Journalism Network. Data Journalism. (2011). Recuperado el 01 de 02 de 2017, de <http://gijn.org/resources/data-journalism/>
- Gómez, P. (2008). Libertad de expresión. Revisión metodológica de las agresiones a periodistas en México. Ciudad de México, México: Fundación para la Libertad de Expresión.
- Haber, J. (2014). MOOCS. Cambridge: Massachusetts Institute of Technology.
- Herederó, C. d., Agius, J. J.-H., Romero, S. M.-R., Salgado, S. M., Navarro, A. M., & Sánchez, J. J. (2006). *Dirección y Gestión de los Sistemas de Información en la Empresa*. Madrid: ESIC.
- Himan, P. (2001). La ética del hacker y el espíritu de la era de la información.
- Himanen, P. (2002). La ética del hacker y el espíritu de la era de la información.
- Houston, B. (8 de March de 2016). Una breve historia de 50 años de periodismo y de datos. Obtenido de *Global Investigative Journal*: <http://gijn.org/2016/03/08/una-breve-historia-de-50-anos-de-periodismo-y-de-datos/>
- Hoyos, R. D. (2006). Disciplinariedad, Interdisciplinariedad, y Transdisciplinariedad. *Vínculos y Límites*. Escritos, 94-125.
- Hunter, J., Dale, D., Firing, E., & Dro, M. (7 de Octubre de 2017). Matplotlib . Recuperado el 1 de Octubre de 2017, de The Matplotlib Development Team: <https://matplotlib.org/>

- Investigative Reporters and Editors. (2015). Coding for Journalist. Recuperado el 18 de Diciembre de 2016, de Coding for Journalist: <https://coding-for-journalists.readthedocs.io/en/latest/>
- Kaplan, D. E. (14 de 01 de 2013). Global Investigative Journalism: Strategies for Support. Recuperado el 16 de 02 de 2017, de Center for International Media Assistance: <http://www.cima.ned.org/wp-content/uploads/2015/02/CIMA-Investigative%20Journalism%20-%20Dave%20Kaplan.pdf>
- Kara, J. (29 de julio de 2016). Real-world Python for data-crunching journalists. Recuperado el 6 de febrero de 2017, de Trend: the story in numbers: <https://trendct.org/2016/07/29/practical-python-pandas-and-jupyter-for-data-journalists/>
- Kearsley, G. (2000). Online Teaching. Canada: Wadsworth.
- Kirschner, P. (2001). Using integrated electronic environments for collaborative teaching/learning. *Research Dialogue in Learning and Instruction*, 1-9.
- Lara, T. (2009). Los medios de comunicación como redes sociales. *TELOS*(79), 1. Recuperado el 21 de 10 de 2017
- Lizana, N. S. (2006). El Periodismo y la Función del Periodista. *Revista Mexicana de Comunicación*. Recuperado el 10 de 11 de 2016, de <http://www.mexicanadecomunicacion.com.mx/Tables/FMB/forouni/periodismo.htm>
- Magallón, R. R. (2014). Tecnologías Cívicas y Participación Ciudadana. *Revista de Estudios de Juventud.*, 105, 53-76.
- Magallón, R. R. (2014). Tecnologías Cívicas y Participación Ciudadana. *Revista de Estudios de Juventud*(105), 53-76. Recuperado el 21 de 10 de 2017
- Manovich, L. (2011). Trending: The promises and the challenges of big social data. *Debates in the digital humanities*, 2, 460-475.
- Manual de Periodismo de Datos. (2012). Recuperado el 20 de 01 de 2017, de La Nación: <http://interactivos.lanacion.com.ar/manual-data/>
- Manual de Periodismo de Datos Iberoamericano. (2015). Obtenido de <http://manual.periodismodedatos.org/>
- Mayer-Shönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. New York: Houghton Mifflin Harcourt.
- Mergel, B. (1998). *Diseño Instruccional y Teoría del Aprendizaje*. Canadá: Universidad de Saskatchewan.
- Meyer, P. (21 de Marzo de 2011). The Difference Between a Web Application and a Desktop Application. Recuperado el 2017 de Septiembre de 2017, de Technology Articles: <http://web-desktop-application.articles.r-tt.com/>
- Miiksenaar, P. (2001). *Una Introducción al Diseño de la Información*. Gustavo Gili.
- Milan, S. (enero-junio de 2015). Medios Ciudadanos y Big Data: La Emergencia del Activismo de Datos. *Mediaciones*(14), 10-26. Recuperado el https://issuu.com/coordinaciondecomunicacionesfccunim/docs/mediaciones_1410 de 11 de 2016
- Milan, S., & Gutiérrez, R. (2015). Medios ciudadanos y big data: La emergencia del activismo de datos. *Mediaciones*, 14, 10-26.

- Milenio Data Lab. (07 de 03 de 2016). Milenio. Recuperado el 24 de 10 de 2017, de Milenio Data Lab: http://www.milenio.com/datalab/falso-exito-Cruzada-Nacional-Hambre_0_696530646.html
- Montfort, N. (2015). El caso de sistemas concisos de creatividad computacional. En R. Pérez, *Creatividad Computacional* (págs. 21-31). México: Patria.
- Montfort, N. (2016). *Exploratory Programming for the Arts and Humanities*. Cambridge: MIT.
- Morín, E. (1998). *El Método IV. Las ideas*.
- Moya Cantero, E. (2011). La ética hacker y el espíritu del informacionalismo. Wikileaks como caso pragmático. *Revista Internacional de Filosofía*, Suplemento 4, 323-332.
- Moya, R. (27 de Mayo de 2015). Scraping en Python (BeautifulSoup), con ejemplos. Recuperado el 22 de Septiembre de 2017, de Jarroba: <https://jarroba.com/scraping-python-beautifulsoup-ejemplos/>
- Mytchell, R. (2015). *Web Scraping with Python: collecting data from the modern web*. Sebastopol: O'Reilly.
- Nikitina, S. (2005). *Pathways of Interdisciplinary Cognition. Cognition and Instruction*.
- Obamayer, B., & Obermaier, F. (2016). *Panama Papers. El club mundial de los evasores de impuestos*. Planeta.
- Obermaier, F., & Obermayer, B. (2016). *Panamá Papers, El club mundial de los evasores de impuestos*. Planeta Mexicana.
- OECD. (20 de Noviembre de 2016). *Estudio de Datos Gubernamentales de México: Resumen ejecutivo*. Paris. Obtenido de OECD Publishing: <http://www.oecd.org/gov/digital-government/DigGovMex-Summary-Spanish.pdf>
- Open Data Foundation. (12 de 01 de 2016). Obtenido de <http://www.opendatafoundation.org/>
- Open School of Journalism. (s.f.). Recuperado el 21 de 01 de 2017, de About Journalism: <https://www.openschoolofjournalism.com/resources/about-journalism>
- Pandas. (7 de Julio de 2017). Python Data Analysis Library. Recuperado el 5 de Octubre de 2017, de Pandas: <http://pandas.pydata.org/#python-data-analysis-library>
- Paoli, A. (1989). *Comunicación e Información. Perspectivas Teóricas*. México: Trillas.
- Parra, D., & Álvarez, J. (2004). *Ciberperiodismo*. Madrid, España: Editorial Síntesis.
- Pavlik, J. (2005). *El periodismo y los nuevos medios de comunicación*. Nueva York, E.U.A: Paidós.
- Pérez Álvarez, R., Maldonado, J., Rendich, R., Pérez Sanagustín, M., & Sapunar, D. (2017). Observatorio MOOC UC: la adopción de MOOCs en la educación superior en América Latina y Europa. *Actas de la Jornada de MOOCs en español en EMOOCs*. Obtenido de <http://ceur-ws.org/Vol-1836/1.pdf>
- Poikola, A., Villum, C., Dietrich, D., Gray, J., Tait, J., Rogers, K., Zijlstra, T. (2009). *The Open Data Handbook*. Obtenido de Open Data Handbook: <http://opendatahandbook.org/>
- Portal de Datos Abiertos de Barcelona. (2016). Obtenido de <http://opendata.bcn.cat/opendata/es>
- Ramírez, F. (11 de enero de 2012). *Periodismo de datos, periodismo de siempre*. Recuperado el 20 de 04 de 2016, de *Periodismo de datos, periodismo de siempre*: <http://www.lapatria.com/blogs/periodismo-de-datos-periodismo-de-siempre>

- Ramonet, I. (2011). *La explosión del Periodismo*. Buenos Aires: Capital Intelectual.
- Raymond, E. S. (2000). *A brief history of hackerdom*. Thyrus Enterprises.
- Raymond, E. S. (2001). *How to become a Hacker*. Thyrus Enterprises.
- Rivadeneira Prada, R. (Enero-Marzo de 1980). *Periodismo, la teoría general de los sistemas y la ciencia de la comunicación*. México: Trillas.
- Rodríguez Brito, A., & García Chico, L. (2013). *Cartografiando el periodismo de datos. Big data y periodismo en el continente americano: cinco casos de estudio*. Revista TELOS, 6-10.
- Rodríguez, E. M. (Julio-Diciembre de 2016). *Panama Papers y el Periodismo en la Red*. *Escritos Académicos de Extensión y Docencia*, 2, 108-119.
- Rodríguez, M. E. (Julio-Diciembre de 2016). *Panamá Papers y el Periodismo en Red*. *Sociales Investiga. Escritos académicos de extensión y docencia*, 2, 108-119.
- Roldán, N., Castillo, M., & Ureste, M. (04 de 09 de 2017). *La Estafa Maestra*. Recuperado el 24 de 10 de 2017, de <https://www.animalpolitico.com/estafa-maestra/>
- Romero, N. (2016). *Manual de Diseño Instruccional: Una Propuesta con Tareas Integradoras*. Digital UNID.
- Ronacher, A. (2014). *Jinja*. Recuperado el 22 de Septiembre de 2017, de Jinja: <http://jinja.pocoo.org/>
- Ronacher, A. (2014). *Werkzeug*. Recuperado el 22 de Septiembre de 2017, de Werkzeug: the python WSGI utility library: <http://werkzeug.pocoo.org/>
- Ronacher, A. (12 de Septiembre de 2017). *Flask Web Development*. Obtenido de Flask Web Development, one drop at time: <http://flask.pocoo.org/>
- Ruiz, A. (. (2006). *Transparencia y rendición de cuentas*. Ciudad de México, México: Distribuciones Fontamara, S.A.
- Sánchez, J. D., Bolaños, M. H., Magallón, R. R., & Anderica, V. C. (2015). *El papel de las tecnologías cívicas en la redefinición de la esfera pública*. *Historia y Comunicación Social*, 2, 483-499.
- Sangrà, A., & Duart, J. (2000). *Aprender en la Virtualidad*. Gedisa.
- Sha, W. X., & Carotti-Sha, G. (2016). *Big Data*. *AI and Society*, 1-4.
- Soares, S. (3 de Junio de 2012). *Not Your Type? Big Data Matchmaker On Five Data Types You Need To Explore Today*. Recuperado el 3 de marzo de 2016, de Dataversity: <http://www.dataversity.net/not-your-type-big-data-matchmaker-on-five-data-types-you-need-to-explore-today/>
- Tamayo, M. T. (1995). *La Interdisciplinariiedad*. Cali: Centro de Recursos para la Enseñanza y el Aprendizaje.
- Tamayo, M. T. (2004). *El Proceso de la Investigación Científica*. México: Limusa.
- Telles, M. (2008). *Python Power: the comprehensive guide*. Boston: Thomson Course Technology.
- Tena-Espinoza-de-los-Monteros, M.-A., & Merlo Vega, J. (2017). *Tecnología cívica para la participación ciudadana. El caso de Codeando México*. *El Profesional de la Información*, 114-124.

- Vaingast, S. (2014). *Beginning Python Visualization: crafting visual transformations scripts*. Apress.
- Valverde Berrocoso, J. (2014). MOOCS: una visión crítica desde las ciencias de la educación. Profesorado. *Revista de Currículum y Formación de Profesorado*, 93-111.
- Villanueva, E. (2004). *Derecho de acceso a la información pública y ética periodística*. Asunción, Paraguay: Instituto de prensa y libertad. Intercontinental Editora.
- Welsh, B. (9 de Julio de 2017). Knight Center for Journalism in the Americas. Recuperado el 10 de Agosto de 2017, de Python for Data Journalists: Analyzing Money in Politics: <http://journalismcourses.org/PDJ0517.html>
- WordPress. (2017). Funciones. Recuperado el 14 de junio de 2017, de WordPress.com: <https://es.wordpress.com/features/>
- Wurman, R. S. (2001). *Angustia informativa*. Pearson Education.
- Younge, G. (5 de 09 de 2011). The Guardian. Recuperado el 08 de 11 de 2016, de The Detroit riots of 1967 hold some lessons for the UK: <https://www.theguardian.com/uk/2011/sep/05/detroit-riots-1967-lessons-uk>

Lista de Ilustraciones

Figura 1. Línea del tiempo de la historia del Periodismo de Datos. (Fuente: Elaboración propia).....	20	Figura 11. Visualización de “Desaparecidos”. (Fuente: http://interactivo.eluniversal.com.mx/desaparecidos/mexico.html Consultado el: 24 de octubre de 2017).....	47
Figura 2. Disciplinas que conforman el Periodismo de Datos. (Fuente: Elaboración propia)).....	22	Figura 12. Equivalencias en bits.....	51
Figura 3. Metodología del Periodismo de Datos. (Fuente: Elaboración propia)).....	25	Figura 13. Clasificación de datos masivos.....	52
Figura 4. Portada del reportaje Chinaleaks. (Fuente: https://www.icij.Org/Blog/2014/01/Highlights-Chinaleaks Consultado el: 21 De Octubre De 2017)).....	33	Figura 14. Elaboración propia basada en la clasificación de Tim Berners-Lee. (Fuente: http://5stardata.info/es/ Consultado el: 15 de octubre de 2017).....	56
Figura 5. Periodistas entrevistando a Carlos Sousa, Dir. de RRPP de Mossack Fonseca & Co. (Fuente: https://Panamapapers.Icij.Org . Consultado el: 7 De Noviembre De 2017).....	34	Figura 15. Página de la sección de datos de la plataforma digital del gobierno de la República Mexicana. (Fuente: https://datos.gob.mx/busca/datase Consultado el: 15 de octubre de 2017).....	58
Figura 6. Twitter del medio ¿Quién Compró?. (Fuente: https://twitter.com/quiencompro Consultado el: 3 de noviembre de 2017)...	42	Figura 16. Ejemplo de la exploración de datos de la plataforma. Disponible en: https://datos.gob.mx	58
Figura 7. Dependencias gubernamentales involucradas. (Fuente: http://www.animalpolitico.com/estafa-maestra/#dependencias Consultado el: 24 de octubre de 2017).....	43	Figura 17. Ejemplo del catálogo de herramientas. Disponible en: https://datos.gob.mx/herramientas	58
Figura 8. Universidades involucradas en el desvío de recursos. (Fuente: https://www.animalpolitico.com/estafa-maestra/#dependencias Consultado el: 24 de octubre de 2017).....	43	Figura 18. Sitio herramienta import.io (Fuente: https://www.import.io/ Consultado el: 21 de octubre de 2017).....	66
Figura 9. Visualización de “El (falso) éxito de la Cruzada Nacional contra el Hambre”. (Fuente: http://www.milenio.com/datalab/falso-exito-cruzada-nacional-hambre_0_696530646.html Consultado el: 24 de octubre de 2017).....	45	Figura 19. Sitio herramienta Open Refine. (Fuente: http://openrefine.org/ Consultado el: 21 de octubre de 2017).....	67
Figura 10. Visualización de los cárteles en los estados a través del tiempo. (Fuente: http://narcodata.animalpolitico.com/ Consultado el: 25 de octubre de 2017).....	46	Figura 20. Sitio herramienta d3js. (Fuente: https://d3js.org/ Consultado el: 21 de octubre de 2017).....	69
		Figura 21. Portada del Manual de Periodismo de Datos. (Fuente: http://interactivos.lanacion.com.ar/manual-data/ : 4 de octubre de 2017).....	70

Figura 22. Home de la página principal del sitio oficial de General Assembly.....	76	Figura 37. Logotipos presentados al usuario. (Fuente: Elaboración propia).....	104
Figura 23. Página inicial del sitio Data Tells a Story.....	77	Figura 38. Nube de palabras de la opinión de los entrevistados (Fuente: Elaboración propia).....	105
Figura 24. Una de las secciones del sitio de Big Data University.....	77	Figura 39. Primer boceto de Dataísta. (Fuente: Elaboración propia).....	106
Figura 25. Sección de cursos del sitio DataCamp.....	78	Figura 40. Prototipo de página de inicio de sesión. (Fuente: Elaboración propia).....	107
Figura 26. Página inicial de DataBasic.....	79	Figura 41. Prototipo del menú de navegación. (Fuente: Elaboración propia).....	107
Figura 27. Plataforma de aprendizaje Tutellus.....	79	Figura 42. Prototipo de foro (Fuente: Elaboración propia).....	107
Figura 28. Interdisciplina aplicada al proyecto. (Fuente: Elaboración propia).....	85	Figura 43. Listado de temas a incorporar en el menú principal del sistema (Fuente: Elaboración propia).....	108
Figura 29. Descripción de las características del usuario. (Fuente: Elaboración propia).....	91	Figura 44. Retícula del sistema. (Fuente: Elaboración propia).....	109
Figura 30. Mapa del Sistema de Información. (Fuente: Elaboración propia).....	98	Figura 45. Diseño del home (Fuente: Elaboración propia).....	109
Figura 31. Diagrama de Casos de Uso – Administrador. (Fuente: Elaboración propia).....	99	Figura 46. Botones del menú principal. (Fuente: Elaboración propia).....	110
Figura 32. Diagrama de Casos de Uso – Usuario Registrado. (Fuente: Elaboración propia).....	100	Figura 47. Menú vertical (Fuente: Elaboración propia).....	110
Figura 33. Diagrama de Casos de Uso – Usuario no Registrado. (Fuente: Elaboración propia).....	100	Figura 48. Acomodo de clasificación de herramientas. (Fuente: Elaboración propia).....	111
Figura 34. Diagrama de Actividad – Ingresar al Sistema. (Fuente: Elaboración propia).....	101	Figura 49. Vista del diseño del sistema. (Fuente: Elaboración propia).....	111
Figura 35. Diagrama de Actividad – Agregar sitio al catálogo de BBDD. (Fuente: Elaboración propia).....	101	Figura 50. Editor de WordPress.....	115
Figura 36. Diagrama de Actividad – Ingresar información al foro. (Fuente: Elaboración propia).....	102	Figura 51. Editor Visual de WordPress.....	116

Figura 52. Editor de Código Fuente de WordPress.....	116	Figura 71. Logos de herramientas auxiliares.....	135
Figura 53. Página de Bienvenida del Sistema de Información, Dataísta.....	118	Figura 72. Resultados de la prueba de usabilidad. (Fuente: Elaboración propia).....	139
Figura 54. Home del Sistema de información, Dataísta.....	119	Figura 73. Resultados de la prueba de usabilidad. (Fuente: Elaboración propia).....	140
Figura 55. Página de Introducción al contenido.....	119	Figura 74. Nube de palabras relacionadas con usabilidad del sistema. (Fuente: Elaboración propia).....	146
Figura 56. Sección de Texto Instruccional y Video.....	120	Figura 75. Nube de palabras relacionadas con el contenido del sistema. (Fuente: Elaboración propia).....	147
Figura 57. Sección de Clasificación de Herramientas.....	120	Figura 76. Aspectos positivos del sistema. (Fuente: Elaboración propia).....	148
Figura 58. Sección de Herramienta Auxiliar.....	120	Figura 77. Focus Groups “Dataísta”. (Fuente: Elaboración propia).....	149
Figura 59. Sección de Medios.....	121		
Figura 60. Sección de Preguntas Frecuentes.....	121		
Figura 61. Sección de Foros.	122		
Figura 62. Sección de Inicio de Sesión y Registro.....	122		
Figura 63. Sección de Datos Abiertos.....	123		
Figura 64. Vista del Clasificador de Herramientas.....	123		
Figura 65. Detalle del Clasificador de Herramientas.....	124		
Figura 66. Etapas de Desarrollo de las Herramientas Auxiliares.....	130		
Figura 67. Interfaz Herramienta Obtención.....	131		
Figura 68. Interfaz de la Herramienta de Limpieza.....	133		
Figura 69. Interfaz de la Herramienta de Análisis.....	133		
Figura 70. Interfaz de la Herramienta de Visualización.....	135		

Lista de tablas

Tabla 1. Ejemplo del catálogo de herramientas. Disponible en https://datos.gob.mx/herramientas	59
Tabla 2. Herramientas auxiliares. (Fuente: Elaboración propia).....	71
Tabla 3. Diferencias entre cursos masivos y cursos en línea. (Fuente: Elaboración propia).....	74
Tabla 4. Personas seleccionadas para entrevistar.....	87
Tabla 5. Temas abordados durante el “Taller de Datos Extremo” (Fuente: Elaboración propia con base en los contenidos planteados por “Escuela de Datos”).....	88
Tabla 6. Resumen de la metodología del proyecto. (Fuente: Elaboración propia).....	90
Tabla 7. Procesos de la Herramienta de Obtención.....	131
Tabla 8. Procesos de la Herramienta de Limpieza.....	133
Tabla 9. Procesos de la Herramienta de Análisis.....	134
Tabla 10. Procesos de la Herramienta de Visualización.....	135

Anexo A: Tecnologías de la Información en el Periodismo

Categoría	Definición	Incluye	Técnicas y Tecnologías
Reporteo de Datos	Obtener, limpiar y analizar datos para contar historias periodísticas.	Reportes asistidos por computadora. El uso de métodos de las ciencias sociales. Mapear y graficar datos para la exploración y el análisis. Programar para obtener y analizar datos.	Acceso a la información pública. Herramientas y técnicas de Web. Scraping. Uso de software de bases de datos relacionales. Entender estadística básica y software especializado. Utilizar herramientas y técnicas de mapeo y visualización.
Visualización de Datos e Interactivos	Se utiliza código para la publicación (HTML, CSS, JS) y bases de datos para construir trabajos periodísticos interactivos.	Desarrollo de visualizaciones y gráficos interactivos. Aplicaciones interactivas para la búsqueda de conjuntos de datos.	El uso de HTML, CSS y JavaScript. El uso de software de visualización como: Tableau y D3.js. Programación y manejo de bases de datos como: Python, Django, Flask y Ruby on Rails. Uso de aplicaciones de mapeo como: QGIS, CartoDB, ESRI, TileMill y GeoDjango. El uso de herramientas para el desarrollo de software.
Emergencia de Tecnologías para el Periodismo	Desarrollar herramientas y productos usando datos y tecnología.	Periodismo de drones. Periodismo de sensores. Periodismo de realidad virtual y aumentada.	Sistemas aéreos para recolectar fotos, video y datos. Sensores para medir la calidad y el nivel de algunas condiciones ambientales. Samsung, OCULUS y CardBoard.
Periodismo Computacional	Se utilizan algoritmos, el aprendizaje automático y otros métodos para apoyar a los objetivos del periodismo.	Algoritmos para ayudar a los periodistas a estructurar datos de nuevas maneras. Nuevas plataformas digitales para manejar mejor documentos y datos.	Lenguajes de programación como: Python, Ruby o R. Marcos y aplicaciones como Jupyter, para permitir al periodista mezclar y analizar código. Plataformas como Overview, que facilitan el uso de procesos complicados del procesamiento del lenguaje natural.

Anexo B: Entrevistas

Tania Lidice Montalvo Solís

Fecha: 20/06/2016

Es editora del portal de noticias Animal Político, además colabora en el proyecto NarcoData, uno de los primeros proyectos de Periodismo de Datos en México. Estuvo al frente del proyecto El Sabueso, primer proyecto mexicano de verificación de datos. Los temas principales de sus investigaciones se relacionan con presupuestos del gobierno federal y subnacionales, corrupción, rendición de cuentas y uso de recursos públicos.

A continuación, se presenta la transcripción de la entrevista realizada.

1.- ¿Cómo definiría a un periodista de datos?

“Bueno, igual y no soy la mejor persona para responder esa pregunta porque yo no creo, como tal que haya un Periodismo de Datos. No importa la clase de reportaje o investigación que vayas a hacer, todo requiere de datos, si no se queda en un reporte de una anécdota sino tiene datos.

De alguna forma, lo que se está llamando ahorita “periodista de datos”, es aquel que basa su investigación en bases grandes, bases de datos; y de ese ordenamiento de alguna forma como que obtienen respuestas, hallazgos o alguna revelación periodística. Digamos que, un reportero puede obtener

la revelación periodística de diferentes formas; entonces, hay veces que quizá puede ser a través de qué es lo que más se usa: una declaración de algún funcionario, hechos específicos o en este caso, datos.

En realidad, se trata de obtener una revelación periodística a través de hechos y finalmente vas a utilizar datos. Pues entonces, un periodista de datos, lo que hace es obtener, sistematizar, interpretar información para obtener alguna revelación periodística o algo de interés público, para convertirlo en periodismo.

En realidad, yo utilizo muchos datos, o sea, yo utilizo por ejemplo Excel o herramientas como... Pues sí, herramientas que permitan utilizar bases de datos, pero en realidad todas las historias que hago, aun cuando sean de campo, alguna historia que sea más anecdótica -más de historias humanas si quieres-, siempre las combino con datos. Y al menos nosotros, en Animal Político todos los reporteros es algo como que tenemos que hacer, o sea, todas las historias deben tener datos; no podemos quedarnos en una anécdota. O sea, hay reportes, por ejemplo, si estás en una manifestación, no vas a poder utilizar datos en específico, porque no los tienes en ese momento y solo vas a reportar lo que estás viendo, datos muy concretos, pero cada historia reportada tendría que tenerlos, pues igual y hay periodistas que su fuen-

te principal son las bases de datos y estos, quizás sean los periodistas de datos, pero en realidad todos los periodistas deberíamos de utilizar datos.”

2. Describa brevemente cada uno de los pasos que lleva a cabo al realizar una investigación basada en datos.

“Pues es que no creo que todo el tiempo sea lo mismo. Cada historia surge de una corazonada periodística, lo que llamamos el olfato periodístico; puede venir de un evento, de una fecha conmemorativa de un hecho en particular. El quehacer como periodista es, pensar que ahí hay una historia escondida, bueno, no necesariamente escondida; sino una historia que puede ser contada. Y entonces, ahí pueden surgir diferentes formas.

O sea, esta historia puede motivarme digamos, a pedir y utilizar la ley de transparencia para pedir información pública y obtener mis primeros datos y, puede ser que utilice otras herramientas; por ejemplo, los datos abiertos que ahorita están de moda y que, muchas veces son útiles, o datos de organizaciones civiles, organismos internacionales, para empezar a sustentar la hipótesis de la historia que quiero contar.

Ya de ahí, puede ser eso primero o incluso empezar ya el reporte. El reporte, como en campo, alguna entrevista, o sea, yo creo que las bases de datos o las

conclusiones con las que llegues con bases de datos, lo ideal es que los confirmes con alguien que sepa del tema, o sea, un experto o alguien que tenga bastante dominio o conocimiento del tema para que confirme tus hallazgos. En realidad, es lo que haríamos con cualquier nota, confirmar la información y bueno, ya de ahí construirla; jugar con los datos, limpiar la base de datos, acomodarla, buscar la forma de que suelte respuestas, bueno, cuestionarla.

Bueno, es un poco literal; lo que haces es preguntarle cosas a los datos. O sea, les pides ciertas hipótesis de la historia que quieres construir y ya que tienes limpia tu base de datos, le haces preguntas a la información y solita te va arrojando respuestas.”

3. Del proceso que describió anteriormente, mencione 3 problemáticas o dificultades que se suscitan durante su desarrollo.

“Pues lo más difícil -al menos en nuestro país-, es obtener los datos, por cierto. Porque, o sea, tenemos la Ley de Transparencia, que ayuda -y ayuda mucho-, pero incluso tenemos, lo que les decía de los datos abiertos, pero en la mayoría de los casos los datos no están listos para ser trabajados; no como debería de ocurrir según los compromisos que hizo el estado mexicano en la Alianza para el Gobierno Abierto; que es entregar los datos abiertos. O sea, a mí me ha pasado que entregan hojas de papel que son fotocopias, o sea, que son impresiones de un documento de Excel o

de un documento que es una base de datos fotocopiada cuatro veces y ni siquiera se entiende. Entonces, ese es el principal obstáculo; obtener información de verdad es muy complicado.

Una vez que la obtienes -por lo mismo que lo normal es obtenerla en formatos que no son nada amigables-, pues hay que arreglarla, hay que acomodarla, hay que buscar la forma de que se convierta en una base de datos. Muchas veces, son PDFs que en realidad son imágenes mal sacadas/chuecas, incluso es muy difícil scrapearlas. Realmente, yo pienso que este es el principal obstáculo porque ya teniendo la información es cuestión de tiempo y dedicación para que encuentres algo.”

4. ¿Qué herramientas utiliza para procesar grandes cantidades de información?

“Lo hago yo -como te decía-, yo siempre tengo hipótesis. Si estoy buscando “desvío de recursos” y siento que hay algo -cuando yo obtengo los datos-, lo que empiezo a buscar -es, no sé-, en los campos sobre dinero o entrega de recursos públicos, empiezo a ver cómo han sido las variaciones porcentuales. Sí ha habido años en donde no se reportan gastos de ese tipo de cosas. Y siempre, cuando yo llego a las bases de datos ya sistematizadas, tengo hipótesis sobre diferentes puntos de vista de la historia que quiero contar. Empiezo a mover los datos, o sea, a mirarlos y moverlos en términos de una suma, una resta, porcentajes, variación porcentual, -te digo lo que sea- tasas para ver los cambios.

De menos, Excel, seguro lo básico, que es una cuestión sencillísima, que pasa -igual y suena sangrón- pero la mayoría de los periodistas no saben sacar un porcentaje en Excel, entonces -este-, si no saben sacar un porcentaje para una base, digamos de cien registros, digamos, la mayoría van a hacer a mano, una por una; la mayoría de los periodistas no saben poner una fórmula y jalarla para que todos sus datos estén así.

Realmente, es una cuestión tan sencilla como un porcentaje; ya no hablemos de contar registros repetidos, de estas, por ejemplo; sí tienes una base de empresas contratistas de 300,000 o 700,000 registros y quieres contar cuántas veces el gobierno le haya dado contratos a la misma empresa; imposible que la mayoría de los periodistas no cuenten con una fórmula en Excel; la mayoría, van a ponerse a contar y eso, obviamente se presta para el error.

Para scrapear si utilizo ciertas herramientas como Tabula o ciertos sitios web que scrapear datos; Refine, ese tipo de cosas y para limpiar la base de datos. Algo de SPSS pero, como no tengo la licencia básicamente lo hago con Excel.”

4.1 ¿Qué características poseen las herramientas que mencionó anteriormente que hicieron que las seleccionará sobre otras?

“En realidad, es porque son las más populares y entonces es de las que ha sido más fácil encontrar la forma de capacitarme o de que alguien me enseñe, básicamente es eso.”

Israel Piña Camacho

Fecha: 22/06/2016

Periodista y coordinador general del medio independiente ¿Quién Compró?, plataforma de Periodismo de Datos para transparentar el uso del dinero público en el Congreso de México.

1. ¿Cómo definiría a un periodista de datos?

“Creo que el periodista de datos es aquel que, echa más de bases de datos en formato abierto para sustentar una investigación, para mostrar evidencias sobre la tesis de su investigación y en general para darle mayor rigor a su investigación.”

2. Describa brevemente cada uno de los pasos que lleva a cabo al realizar una investigación basada en datos.

“Para empezar, hay que hacer una hipótesis de trabajo; planteas un problema -el principal- y, enseguida una hipótesis sobre ese tema. Una vez que tienes la hipótesis, busca la información y la manera de hallar esa información -puede ser por diferentes vías-, puede ser una búsqueda en la web, puede ser un scraping, una minería de datos más profunda, pueden ser solicitudes vía transparencia, en fin...”

...Después de eso, lo importante es manejar la información y eso empieza por limpiar tu base de datos. Generalmente las bases de datos vienen con errores, vienen sucias, o simplemente no todo lo que viene en una base de datos te sirve para el

objetivo de tu investigación; lo que tienes que hacer es limpiar la base de datos, analizar los datos en sí. Una vez que analizas los datos, encuentras donde está como lo anormal en la distribución, lo que rompe con la regla, la desviación y entonces ahí hay una historia.

Analizas los datos para encontrar una historia y cuando encuentras la historia, entonces sales a la calle. Ser periodista de datos, no significa hacer todo desde tu escritorio y ordenador; una vez que localizas en los datos una posible historia, sales a la calle, reportas, investigas como siempre se ha hecho el periodismo de investigación y después de eso, regresas, visualizas tus datos y los cuentas junto con la historia que ya traes, esos serían los pasos.”

3. Del proceso que describió anteriormente, mencione 3 problemáticas o dificultades que se suscitan durante su desarrollo.

“En nuestro caso particular de ¿Quién Compró?, el problema más importante es que, el Congreso, no tiene bases de datos abiertos; número 1, número 2; las construimos nosotros a través de las respuestas que recibimos vía transparencia. Pero ahí, también hay un problema, porque las Unidades de Enlace -como se le llama a las unidades de transparencia en el Congreso-, prácticamente lo que hacen es “bloquearte”; su función no es darte la información, sino hacer todo lo posible para no darte la información. Entonces, es difícil conseguir

la información que tú quieres, tienes que estar dando muchas vueltas, utilizar el lenguaje o la frase correcta, pelearte un poco -porque son muy reticentes a darte información-; ese es el segundo problema que enfrentamos.

Y, el tercer problema, es que a veces, la información no existe; no nada más no está en bases de datos o es difícil que nos la den, sino que simplemente no existe porque la ley no obliga a los legisladores a registrar todo. Un ejemplo, hay una partida tanto en diputados como en senadores que se llama “subvenciones”, esas subvenciones es dinero que le dan a las bancadas, no tienen la obligación de transparentarlas con la copia o la factura; simplemente hacen un reporte trimestral, pero no hay un respaldo documental que demuestre cómo se gastaron las subvenciones. Entonces, evidentemente no se puede construir una base de datos con algo que no existe; tampoco te lo pueden dar con transparencia porque no existen y legalmente ellos pueden no transparentarlo, no están obligados, por lo tanto, estas limitaciones legales, serían el tercer problema.”

3.1 ¿El congreso te entrega esa información como fotocopias?

“Sí, no es información abierta, no es información en un formato abierto, son fotocopias y por lo tanto nosotros tenemos que convertir eso en una base de datos. No nos dan bases de datos, mucho menos en formato abierto “.

4. ¿Qué herramientas utiliza para procesar grandes cantidades de información?

“Hay diversas herramientas que se utilizan para cada etapa del proceso; por ejemplo, para visualizar a mí me gusta Tableau, me parece que es la más potente.

Para sistematizar datos, manejar la base de datos usamos Open Refine, y para hacernos llegar información para nosotros lo más importante es vía transparencia, pero cuando hacemos búsquedas, utilizamos desde las herramientas avanzadas de Google, hasta algunas herramientas para capturar tablas en HTML y volverlas una base de datos abiertas. Por ejemplo, Table Captor es una, Import es otra; y luego, hay herramientas para convertir PDFs a base de datos. La más importante que hicieron en La Nación en Argentina y se llama Tabula, lo que hace es, convertir PDFs en tablas.

En general son esas las que usamos, por ejemplo, para visualizar hay unas más sencillas como Piktochart, pero nosotros usamos esas. Hay una nueva, que todavía no la usa todo el mundo porque tienes que pedir permiso para que te dejen usarla; es una que se llama Atlas, es de visualización y la hizo un medio que se llama Quartz.”

4.1 ¿Todo lo van llenando en Excel y de Excel llegan a trabajar con estas otras herramientas?

“A veces, pero también a veces directamente en las herramientas.”

4.2 ¿Tableau, es sencillo de manejar?

“Si, necesitas de una cierta curva de aprendizaje, algo de práctica, no es tan sencilla como Piktochart o como Excel, si tiene un mayor grado de complejidad, pero tampoco es cosa del otro mundo.”

4.3 En estas que mencionaste que utilizas, ¿necesitas algún conocimiento como en lenguaje de programación o incluyen tutoriales; están en español o en inglés?

“Están en inglés; hay tutoriales y no necesitas saber programar. Hay un lenguaje de programación para hacer visualizaciones que se llama D3. Utilizas D3 con JavaScript, pero es muy avanzado, necesitas más tiempo; cuando tienes un equipo pequeño y no tienes programador lo mejor es utilizar una de estas herramientas que no te exigen saber programar. Lo que sí tienes que saber son algunas funciones para manejar hojas de cálculo, en el caso del procesamiento de los datos; funciones más básicas, pero más precisas.”

5. ¿Qué características poseen las herramientas que mencionó anteriormente que hicieron que las seleccionará sobre otras?

“Porque son gratuitas. Tableau, tiene una versión de paga; pero también hay una versión gratuita. Son gratuitas, no implican que sepas programación y, sobre todo, que son muy potentes -son muy buenas-.”

Irving Omar Morales Agiss

Fecha: 24/06/2016

Doctor en Física Nuclear por la Universidad Nacional Autónoma de México. Profesor en la Facultad de Ciencias Complejas de la misma universidad. Miembro fundador del despacho de visualización de datos Morlan y del primer portal de periodismo de datos en México La Data MX.

1. ¿Cómo definiría a un periodista de datos?

“Para mí, es muy difícil definir a un periodista de datos. Una, no soy periodista y dos, pues he trabajado con datos desde hace 15 años; prácticamente toda mi vida. Soy científico y la primera vez que escuché el concepto de periodista de datos, obviamente me vino a la mente que no todo el periodismo es de datos; que no todo el periodismo de datos, que no todo lo basan en datos. Porque para un científico es muy claro y muy fácil pensar que todo debe estar basado en datos, nos entrenan prácticamente para todo el tiempo estar corroborando, basado en datos. Entonces, la primera vez que yo escuché el Periodismo de Datos dije, de qué están hablando, ¿no? Poco a poco, cuando me empecé a involucrar me pareció que la idea del Periodismo de Datos, pues tenía que ver más con utilizar bases de datos, como si fueran personas; como si fueran fuentes de información, y desde ese punto de vista la labor

del periodista de datos, es una labor que no cualquiera puede hacer, ¿no?

Un periodista de datos, de alguna manera es alguien que le gusta meterse a la base de datos, meterse a listas enormes de números y empezar a buscar patrones, ¿no? Hay periodistas que son expertos en buscar patrones en la gente o en los comportamientos de la gente, o en los decires de la gente y, hay otros periodistas que no les gusta tanto eso y que les gusta mucho más ir y meterse en tablas inmersas de datos. Para un científico, es muy raro encontrar un periodista que le guste eso, pero pues está muy bien, creo que en la actualidad es necesario que cada vez haya más periodistas de datos, ¿no?

Un periodista de datos, le gusta andar buscando datos, le gusta andar buscando cosas por internet, ¿no?, conozco muy buenos periodistas de datos que les gustan las dos partes. Le gusta la parte de ir y reportar, y también les gusta buscar datos en internet. Un periodista de datos, también debe tener paciencia para checar listas y listas interminables de datos. Estar limpiando bases de datos, estar corroborando información también debe tener un poco de intuición acerca de cómo presentar estos datos.

Cuando uno, empieza a trabajar con bases de datos que ya son mucho más grandes, bases de datos que son masivas -de alguna manera-, pues ya no es la manera de presentar los patrones que uno encuentra; los resultados o los hallazgos.

Ya no es necesariamente, el periodismo tradicional, ¿no?, los patrones que uno encuentra en una base de datos no necesariamente, ¡eh!, lo más fácil es ponerlos en un párrafo de texto. Entonces, también es muy importante la parte de visualización de datos, de cómo presentar la información de este tipo de formatos, entonces un poco de todo eso.”

2. Describa brevemente cada uno de los pasos que lleva a cabo al realizar una investigación basada en datos.

“Como les decía, yo soy científico; analizo datos desde hace prácticamente 15 o 20 años, ¿no?, y entonces...

...Pues a mí, me gusta mucho estar acumulando datos, de hecho, si les pudiera mostrar mi computadora verían que tengo una carpeta enorme en donde todas las bases que he analizado, las guardo -y casi casi que las colecciono-, y eso me sirve mucho para después, en futuras investigaciones, empezar a cruzar con bases de datos que tenía antiquísimas, ¿no? Las bases de datos que utilizo, la principal fuente yo creo que sería INEGI, ¿no?, conozco prácticamente, muchas de las encuestas de las bases de datos que tiene INEGI, y también bases de datos abiertos del gobierno; me gustan también las bases de datos que construyen por su propia cuenta. Como científico, también me gusta que vienen de sensores, otro tipo de cosas, de hecho; por ejemplo, imágenes de cámaras y de ahí, sacar datos, todo ese tipo de cosas.”

3. Del proceso que describió anteriormente, mencione 3 problemáticas o dificultades que se suscitan durante su desarrollo.

“Sí, la visualización de datos es una herramienta muy poderosa para presentar la información; pero usualmente el periodismo de datos, tiene que ir -desde mi punto de vista- en total acuerdo con lo que dice el periodista y eso, es uno de los principales problemas que he encontrado en las redacciones, que es que, el periodista de datos escribe su nota por su cuenta y, el visualizador de datos hace su visualización, y luego las pega -literalmente- como si fuera un chilaquil y entonces, muchas veces uno lee la nota, y entiende perfectamente lo que quiere decir el periodista. Luego y vas a la visualización de datos y dices “¡ah, qué bonito, aquí están los datos!”, pero uno nunca encuentra la relación entre los dos productos; uno nunca encuentra una narrativa donde, el periodista utilice la visualización de datos para contar una historia.

Creo que, apenas va empezando todo esto del Periodismo de Datos con visualización de datos, y creo que poco a poco, se empezarán a desarrollar técnicas narrativas que utilicen ambas partes, por el momento ese es el principal problema.”

4. ¿Qué herramientas utiliza para procesar grandes cantidades de información?

“Por mi trabajo anterior, mi trabajo usual es ser científico. Entonces, en la academia, utilizamos software mucho más especializado

que el que usualmente se utiliza en el Periodismo de Datos.

Yo, por tradición utilizo un software que se llama MATHEMATICA, que literalmente es un monstruo. Puedes hacer - jeh! - prácticamente cualquier cosa que tenga que ver con análisis o con matemáticas. Pero hay gente que utiliza R, utiliza SPSS; entonces, se vale cualquiera, cualquiera en el que te sientas cómodo y que permita ser el análisis que tú estás buscando. Al final de cuentas, analizar o hacer Periodismo de Datos o simplemente, buscar patrones en los datos, entonces con la herramienta que te sientas mejor, esa es la adecuada.”

5. ¿Qué características poseen las herramientas que mencionó anteriormente que hicieron que las seleccionará sobre otras?

“Bueno, como les decía, yo más que nada; yo, por estar acostumbrado a utilizar MATHEMATICA, ¿no? Para mí, por ejemplo, hace unos meses traté de aprender R, y se me hizo interesante; pero no se me hizo nada del otro mundo porque, porque yo ya estoy muy viciado con MATHEMATICA. Entonces, este pues, esencialmente uso MATHEMATICA, porque me permite hacer lo que quiero hacer, sé hacerlo muy rápidamente, ahí porque llevo quince años usándolo, pero este prácticamente cualquiera puede utilizar cualquiera de estas herramientas.

Para visualización; ese es otro tema por completo. Ahorita hay muchas herramien-

tas para hacer visualizaciones rápidas, a mí no me gustan tanto; prefiero tener un acercamiento a las visualizaciones más a artesanal, digamos. Entonces, todo lo que hacemos nosotros, lo hacemos en una librería que se llama D3.js de Javascript, porque, bueno; es difícil de usarla y cada visualización requiere de un proceso prácticamente artesanal de ir escribiendo código, e ir armando cada punto y cada línea. Pero, es una de las pocas herramientas que te permite, prácticamente lo que tú te imaginas, hacerlo en pantalla.

Si uno de repente, escoge herramientas que son mucho más rápidas y en las redacciones, gran parte del tiempo tratan de utilizar estas herramientas, porque es mucho más rápido hacerlo -mucho más fácil-, pues lo que uno obtiene son, visualizaciones que de alguna manera se ven genéricas. Que de alguna manera están ya, en un montón de lugares y todos tienen más o menos el mismo aspecto; a mí no me gusta eso, porque de alguna manera es como nada más “agarré esta gráfica y la pegué ahí”. No, a mí me gusta que sea una visualización mucho más artesanal, mucho más cuidada -digamos-.

Lilia Saúl Rodríguez

Fecha: 11/08/2016

Lilia Saúl, periodista mexicana especializada en temas de transparencia y acceso a la información, ganadora del premio Ortega y Gasset en 2016. Fue editora en El Universal Data y actualmente se encuentra trabajando en CONNECTAS; donde promueve y trabaja en la producción y difusión de información en temas sobre América Latina también es parte del comité de la comunidad Chicas Poderosas México.

1. ¿Cómo definirías a un periodista de datos?

“Bueno, es complejo porque, en realidad un periodista de datos tendría que tener ciertas habilidades que, en ocasiones un periodista que -que ahora sí que- de día a día no tendría, quizá, ¿no? Una de ellas es quizás, tener más experiencia en cómo usar o cómo visualizar los datos, cómo ordenar los datos primero.

Estamos hablando de que, pudiera tener posibilidades de cómo manejar, por ejemplo, herramientas como Excel, organizar o sistematizar bases de datos. Otro tipo de herramientas, por ejemplo, estoy pensando en estas que ya son como más complejas, pero bueno, sería deseable que pudiera hacer algo con Strata o en R, pero al mismo tiempo algo básico.

Creo yo que un periodista de datos tenía que haber sido antes un periodista nato -de a pie-, que pudiera saber investigar a priori,

porque sabe perfectamente dónde buscar los datos, cómo investigar al mismo tiempo, cómo buscar las historias y cómo saber en dónde encontrar la información; porque muchas veces lo que me he dado cuenta es que la gente que se dice como periodista de datos, no sabe ni siquiera encontrar los documentos o los datos, o cómo puede allegarse de información que es necesaria o útil para su reportaje o su investigación.

Entonces es; uno, es saber utilizar las herramientas con las que cuentas a tu alcance, porque puede ser manejo de bases de datos, saber buscar la información y quizás tener no el conocimiento del todo, pero sí la habilidad para saber cuánto tiempo o qué herramientas para visualizar podrías utilizar para tu trabajo, ¿no? Porque luego -a veces-, también tienes todas las herramientas, todas las bases de datos del mundo, pero a la hora de querer estructurar la información, pues no sabes cómo y te cuesta mucho trabajo. Cómo transmitir ese mensaje o esa ilustración que tú quieres y sino no lo sabes hacer, pues, aunque tengas una maravillosa base de datos que te pueda dar una gran investigación, no sirve de mucho.”

2. ¿Te consideras periodista de datos?

“Sí y no. Sí, porque -antes que te diga sí-, pero antes que nada yo me considero una periodista que empezó, pues ahora sí que de a pie. En muchas de las cuestiones, de cubrir todo tipo de fuentes, yo tenía conocimientos previos de muchos tipos de

información de manera distinta para obtenerla; ya sea, desde solicitudes de información hasta el reporte. Este darle seguimiento a ciertos temas, de ciertas noticias. Y periodista da datos, sí, quizás yo tengo una carencia o un déficit de que no soy una experta en análisis de datos; sí, tengo todavía que estar batallando un poco con Excel, hacer -como que digamos-, bajar la base de datos y luego convertirla, etc., pero creo que tengo un poco eso, tengo -como que digamos-, las otras partes que te digo, que es el hacer uso de las herramientas, el buscar dónde encontrar la información y el cómo saber visualizar, porque creo que eso es algo en mi caso muy particular, me ha llevado a saber con quién juntarme, con quién buscar cómo visualizar los datos o qué queremos hacer, -yo creo-.”

3. ¿Nos podrías describir el proceso que llevas a cabo para realizar una investigación basada en datos?

“Sí, son diferentes procesos. Pero creo que varios de ellos tienen algo similar que es esto que te estoy contando. O sea, primero -hacemos como-, nos planteamos varias preguntas para llegar a algo que queremos revelar, descubrir o investigar; por ejemplo, un día -estando-, buscando un cajero automático, y querer saber varias variables entre ellas, que no te asalten, tener dónde estacionar el vehículo, que sea un banco que no me vaya a cobrar comisión porque no sea de mi red, etc. Todas esas preguntas, aunque parezcan muy obvias o muy

comunes, te llevan a querer pensar en alguna investigación de un reportaje ¿cuánto están recibiendo los bancos por comisiones?, -cuántos- en los bancos ¿cuántos cajeros automáticos hay?, ¿son -los mismo-, el mismo número de cajeros automáticos que tendría este banco en España?, ¿invertirían lo mismo aquí que en otros países o son muy pocos para la gente que somos?

Cuando son días de quincena, ustedes se han dado cuenta de las filas enormes, que tu vas a un cajero y entonces, todo empieza con ese tipo de preguntas que estamos pensando -bueno-, ya la parte quizás última de todas esas preguntas, es quizás, sí, pero ¿cómo consigo la información, - ¿no? -. Entonces, en varias ocasiones, cuando nos hacemos este tipo de preguntas, nos damos cuenta de que no hay las bases de datos que nosotros quisiéramos encontrar.

Entonces, al hacer precisamente este tipo de preguntas con relación al tema los cajeros automáticos, te das cuenta que los bancos no rinden cuentas en ese aspecto. No hay información pública que te pueda permitir saber si lo que invirtió un banco en México, es lo mismo que invirtió en infraestructura en otros países. Tendríamos que hacer nuestra propia base de datos, armarla, saber cuántos cajeros automáticos hay en nuestra ciudad y luego entonces, comenzar la siguiente fase de la investigación que sería hacerle preguntas a esas bases de datos; -pero bueno- entonces, primero empezamos con esas preguntas, los temas que queremos investigar; después, buscamos la

información, o sea, donde podemos encontrar los datos si son fáciles de encontrar. Si son difíciles de encontrar, si los tenemos que armar nosotros mismos, por ejemplo, el caso de las desaparecidas. Cuando se planteó ese proyecto dijimos “bueno, está muy padre, pero es un proyecto que va a tardar meses porque la base de datos la tenemos que construir de manera manual -nosotros-, no hay manera de que tú puedas encontrar esa base de datos porque, aunque están los datos públicos, los datos no están de tal forma que te permita bajar la base de datos en un día, sino que tienes que hacer la lista de uno por uno.

Entonces recapitulamos, primero hacemos las preguntas, después buscar cómo vamos a encontrar los datos, cómo vamos a encontrar las bases de datos; después hacerles las preguntas conducentes a las bases de datos y a la par empezar con la reportería o comenzar con las historias que tienes que reportear en detalle. Porque eso es algo también muy importante; no olvidar que cualquier proyecto que nosotros hacemos, siempre empezamos con las preguntas, pero también pensar ¿qué necesitamos a la hora de hacer la reportería?, ¿a quiénes vamos a acudir?, ¿qué historias queremos contar?, porque siempre, detrás de cualquier número hay una persona de carne y hueso. Entonces, ya que empezamos con esos dos carriles al mismo tiempo, que es, hacer las preguntas a las bases de datos o ir analizando las bases de datos, hacer las historias y ya, después de eso

comenzamos también a pensar cómo queremos visualizar esa información, qué queremos hacer, si solamente serviría hacer una infografía estática en la versión impresa del periódico o si vale la pena también hacer un desarrollo interactivo multimedia. Si es un desarrollo multimedia, va a traer solamente video o foto y a lo mejor algunas gráficas, o va a traer algo más complejo, mapas –por ejemplo- ¿cómo se van a desarrollar esos mapas o si podemos ir a un siguiente nivel, podría ser incluso una aplicación para los celulares. Una app y que también estamos pensando hacer a futuro, si va a ser un proyecto que se va a morir al instante de ser publicado; -me refiero-, se publica, se da un golpe mediático y se queda ahí, o si se va a dar continuidad al proyecto; o si va a ser un proyecto que puede vivir todo el tiempo que se quiera, de manera constante porque es información que se va a estar actualizando constantemente.

Entonces, todas estas etapas que implican una serie de personas que obviamente no solamente soy yo. Implica, por ejemplo, la reportería que puede ser alguien que esté conmigo en el área, implica el análisis de datos que, para eso está, por ejemplo, mi colega Daniela Guazo; que aparte de ser periodista, ella también hace un buen análisis de la información de los datos, de las bases de datos. Implica que estemos trabajando por ejemplo con diseñadores web para hacer las maquetas hasta ver qué herramienta podemos usar para visualizar este interactivo. Infografistas, que son los

que están haciendo el trabajo ya en el periódico, para al mismo tiempo trabajar con nosotros para hacer estas infografías en la versión impresa; y ya por último, también planear una estrategia en redes que también, algo que -también- se deja de lado, ya cuando estás de este lado del periodismo, nos hemos dado cuenta que si hay algo que ha dado, digamos, muy buenos resultados es que no solamente tu investigación puede ser buena, pero no solamente necesita ser buena, sino que además la gente sepa que es buena. Y para que la gente sepa que es buena, la gente necesita conocer el proyecto, ¿y cómo lo van a conocer?, si no te acercas a las redes sociales que son finalmente los aliados fundamentales del periodismo y también del Periodismo de Datos.”

3.1 ¿De que depende la parte visual, siempre hay alguien trabajando está parte como el infografista?

“Es que depende de muchas cosas. Depende de, por ejemplo, que sea una base de datos que a lo mejor no tenga la complejidad o que no tardemos muchos meses en poder, por ejemplo, llevar a cabo un multimedia. Porque, creo yo también que hay ciertas investigaciones que no puedes tardarte un poco más de seis meses, entonces ya cuando lo publiques, quizás la coyuntura ya pasó o a lo mejor ya está muy rebasado. O algo que por ejemplo nos pasó, -que no está mal tampoco, cada país tiene su propia complejidad- nosotros sacamos un proyecto que se llama Viajes Transparentes,

que se basó en solicitudes de información de todos los viajes al extranjero que han hecho todos los funcionarios públicos, sobre todos los Secretarios de Estado fuera del país y pedimos, facturas, pedimos boletos de avión, pedimos justificaciones de que fueron a hacer fuera del país, y nosotros estábamos sistematizando la información cuando Ojo Público, el sitio web de Perú, hizo prácticamente lo mismo, pero en el caso solamente del Congreso. Y lo vimos y dijimos “¡Wow!”, es nuestra idea, el problema fue que nosotros nos tardamos más tiempo, de entrada, porque las solicitudes de información no caen tan rápido como nosotros quisiéramos, porque hubo información que se tuvo que pagar, y la información, la mayoría de ella se tuvo que sistematizar de manera manual porque nos entregaron todo en PDFs y en papel. No nos lo mandaron por correo electrónico o por el sistema Infomex, entonces ahí bueno, te das cuenta de que no estuvo mal pensar en el proyecto de esa manera, de considerarlo para un multimedia, pero sí creo que hubiéramos podido tener como posibilidades de sacar el proyecto más rápido si hubiéramos tenido el acompañamiento desde un inicio con un programador y el diseñador web, que nos hubiera ayudado a saber cómo organizar la información de manera más rápida para que no costara tanto trabajo para nosotros y, para sacar más rápido el multimedia, porque en realidad nosotros teníamos prácticamente la reportería y la investigación, lo que nos

hacía falta era cómo visualizar el trabajo, y cuándo vimos que eso lo sacó Ojo Público, dijimos “¡Oops!”, creo que nos está haciendo falta ya, tener el acompañamiento del programador, pero pues no lo habíamos hecho porque nosotros todavía teníamos que entregárselo hasta que acabáramos de sistematizar toda la información, digo, son de esos aprendizajes que te das cuenta –bueno a ver-, al preguntarte cuánto tendríamos que tardarnos para tantos proyectos que son con tan pocos datos, que solamente los hacemos para la versión impresa, que era lo que -ayer- justo lo que le estaba comentando ayer al infografista; deberíamos de tener al menos una cuenta en Twitter, para mostrar todos los gráficos que has hecho para la versión impresa del periódico. Porque casi no se ven, o sea, si tu no compras el periódico y no lo ves –es muy difícil- en tú red social de Twitter o Facebook, conozcas las infografías que hemos hecho con el área de infografías, por la versión impresa del periódico, casi no suben ese material, por cuestiones quiero pensar que por falta a lo mejor de pericia o de interés, -qué se yo-, pero el trabajo que hacen los infografistas. No se ve, no se está viendo, entonces creo que es una carencia que tenemos.

Pero bueno, regresando al tema de cuánto tiempo decidimos, pues sí lo decidimos en equipo, pero actualmente en el equipo solo está Daniela y yo, entonces es muy fácil resolver muchas cosas, precisamente porque somos pocas.

Somos ella y yo, necesitamos pensar cómo lo vamos a hacer, para que en determinado tiempo podamos decir, a qué le vamos a dar prioridad, - ¿no? -; por ejemplo, ahorita estamos trabajando en un proyecto que sabemos no tenemos prisa en sacarlo porque es un proyecto a largo plazo y podemos tardarnos los seis meses; porque es un proyecto que lo estamos haciendo entre ella y yo, y ya acordamos tener varias reuniones de manera consecutiva con un diseñador web y con alguien del equipo de infografía del periódico, entonces no hay problema, podemos tardarnos seis meses sin que nadie nos esté carrerando. En cambio, hay otros trabajos que tenemos que, -digamos, suena feo decir que- tenemos que cumplir con la cuota, -digamos así-, pero pues para que la gente no deje precisamente de saber que existimos en el Periodismo de Datos, pues incluso son trabajos que solamente se publican en la versión impresa, que salen en la página web pero que no traen interactivo, por ejemplo.”

4. ¿Cuál sería el reto más importante a nivel de datos?

“Pues, el reto en México, sobre todo es que no hay datos; a ver sí hay datos, hay muchos datos, pero datos viejos, datos que periodísticamente hablando no sirven de mucho, y que, para temas importantes o coyunturales, para ser una investigación periodística a profundidad. A veces, son difíciles de encontrar, o sea, me refiero, ahorita tenemos un buen; tenemos por un

lado excelentes bases de datos con relación a homicidios, de inseguridad. Te iba a decir desaparecidos, pero no es correcto; porque, porque hasta en el tema de desaparecidos las bases de datos todavía tienen muchas deficiencias; porque sí estará en una base de datos pública y abierta, pero tiene muchas carencias y no es el número real de personas desaparecidas que hay en este país. Debería de haber una base de datos que te dejara satisfecho en el sentido de que dijeras, “esta base de datos está completa y me sirve”, pero no, la realidad es que no es así.

Entonces, uno de los grandes retos que tenemos es, precisamente cómo allegarnos de información de bases de datos; es increíble que, por ejemplo, que para hacer un mapeo de la deuda externa en México, utilicemos las bases de datos del Banco Mundial, porque las de Banxico están pésimas, no puedes descargar un solo repositorio, toda la deuda externa de México, tienes que darle clic por clic a cada uno de los años o de los meses -ya no recuerdo-, y eso es una tontería, no tendrías por qué hacerlo de esa manera; en cambio en el Banco Mundial, tú das clic a cierta cantidad de años que tiene toda la información y ya la tienes, - ¿no? -, exportaciones e importaciones de México, está la información. Mejor la información en el Atlas del MIT, que no sé, la de la Secretaría, en la de ProMéxico, -no sé, estoy pensando algo así que podría ser-, pero en temas delicados y que son temas importantes para nosotros, para

cualquier persona de a pie, no hay datos.

Hay un trabajo que quisimos hacer replicando un símil de una organización; de una organización de Estados Unidos en Florida, que es saber, que restaurantes están sancionados por sanidad y, para no acudir a este tipo de restaurantes y hacer el mapeo y decirte a ti, ciudadano “a este restaurante no vayas, porque aquí se encontraron cucarachas, en este otro se enfermó la gente”, no hay datos. Esos reportes que tiene que hacer por solicitud de información que te dicen, no tenemos esos datos, o está información que les digo de los gastos; algo tan fácil, tan sencillo, que debería estar de manera pública no está; entonces, tuvimos que hacer solicitud de información; dependencia por dependencia. En el caso de las sanciones a los partidos políticos que a todos nos interesa saber cuánto están quitándoles de dinero, para luego entonces, no hay manera; tengo que hacer una solicitud de información porque a pesar de que se les obliga a los partidos a que esa información sea pública, no está. Y así, te puedo decir de cualquier tipo de tema; porque ahorita están las construcciones; nos estamos quejando todos porque las construcciones nos orillan a estar desviándonos, toda esa información pública y son bases de datos que a nosotros nos servirían mucho. Para decirte, por ejemplo, al ciudadano común por dónde no cruzar la calle, o si saber que su propiedad va a devaluarse, porque al día siguiente le van a construir al lado un penal; miles de variables que implican tomas de

decisiones para ti, que podrían servir por el servicio público, y no lo estamos logrando porque no hay información, entonces ese es el primer gran reto.

Ahora, cuando hay datos, ¿a qué nos enfrentamos? La mayoría vienen sucias, es decir -a que me refiero, no hay forma de que tú puedas confiar cien por ciento en esa base de datos que te están entregando. Simplemente, hicimos un ejercicio hace dos años de la UNAM, para saber el ingreso de los estudiantes, de qué escuelas provenían y todas las celdas de la información de las escuelas, aunque eran las mismas escuelas, estaban escritas de manera diferente; en lugar de que dijera Universidad del Valle de México, en todas las celdas decía UVM, U. del Valle de México, entonces llega un momento en el que se tiene que organizar las bases de datos. Porque no puedes trabajar así, porque cuando lo lleves a un programador te va a decir “es que esto está muy sucio, no puedes hacer una lectura, no puedo programar nada” entonces, de entrada, pues todas las bases de datos, no puedes confirmar en ellas, tienes que estandarizarlas, normalizarlas, limpiarlas, buscar que los formatos no se te desconfiguren a la hora de pasarlas con un programador.

Eso es todo un reto, o entenderlas, porque si hay algunas que si puedes confiar en los datos, otras no, entonces ahí nos ves hablando por teléfono con el fulanita que hizo la base de datos, para preguntarle cómo diablos se le ocurrió ponerle un menos a una columna que dice que es positivo,

- ¿no? -, entonces, si son como complejidades de todos los días y pues, básicamente eso.

Y el otro reto tremendo creo, ni los mismos dueños de las bases de datos se responsabilizan de sus bases de datos.

A qué me refiero, en la Secretaría de la Función Pública, hicimos un trabajo sobre los funcionarios sancionados; la base de datos que nos entregaron estaba bien, tenía estructura, te daba incluso datos personales –de cierta manera-, porque traía el nombre y apellido del funcionario en cuestión, que había sido sancionado. Y a la hora de hacer la entrevista con el secretario de la Función Pública, -bueno-, con el subsecretario, nos dijo datos diferentes, y cuando le mostramos la base de datos que ellos mismo nos habían entregado, se quedó, “quien sabe quién les dio esa base de datos, pero eso no coincide con los datos que yo tengo”, obviamente nosotros nos quedamos con la base de datos que nos habían dado, porque ya teníamos procesados los datos y lo que nos estaba diciendo ese secretario, nos tronaba todo.

Otro es que, a veces, ni ellos mismos saben la información que tienen; entonces, ese es un gran reto a la hora de sacar esas investigaciones; porque te das cuenta que no ellos mismos saben la información que tienen y después de publicar la información, los datos que ahí están, son públicos y que te dicen lo que ellos están queriendo negar, entonces es muy complicado.”

5. ¿Para realizar tus investigaciones que tipos de datos utilizas, que fuentes?

“Pues es que hay varias fuentes, de estas que les he estado platicando, - ¿no? -. Desde hacer solicitudes de información, que ahora ya -para nosotros- es algo muy natural estar haciendo constantemente solicitudes de información. La otra es buscar bases de datos públicas, tanto nacionales como internacionales; a qué me refiero, nacionales pues el sitio de INEGI es maravilloso. Incluso, por ejemplo, hay bases de datos que ya puedes buscar por la Auditoría Nacional de la Federación, de otros organismos públicos como, el mismo CIDE, por ejemplo, ente muchos otros.

Ya nivel internacional, pues muchos otros que les comentaba, del Banco Mundial, del Fondo Monetario Internacional, hay incluso una organización que, para luchar contra el lavado de dinero, que también tiene bases de datos; y, por último, que yo creo son las más difíciles de obtener, pero al mismo tiempo que las que más satisfacción nos han dado son las bases de datos propias. O sea, las que tú construyes, por ejemplo, a partir de un monitoreo de notas periódicas, por ejemplo, o de sistematizar información que ya está es pública, pero que nadie más la ha sistematizado. Porque una, porque no la quieren sistematizar, y dos, porque a nadie se le ha ocurrido sistematizarla. Me refiero al trabajo de diputados y senadores de la República, del cargo de los diputados y senadores de la República, del cargo de los diputados, de los senadores,

pero no está de ninguna manera sistematizada en dato abierto para que les puedas hacer análisis de esa información, entonces es donde nosotros sacamos los datos.”

6. ¿Podríamos decir, que realmente hay datos abiertos?

“No, es una falacia. Quisiera decir que sí hay; pero todavía nos falta mucho. De entrada, creo que hay una falta de conocimiento, cuando estamos hablando de datos abiertos, incluso quiero referirme al tema de las estrellas cuando estamos hablando, Tim Berners-Lee, estamos hablando de diferentes niveles de datos. Entonces, cuando tú ya puedes hacer incluso un cruce con toda la información que tú tienes, como si fuera un caleidoscopio.

Esos mismos datos, te pueden dar datos diferentes de la misma base de datos; entonces, eso sí sería maravilloso, pero eso ahorita en México no es posible, incluso podría pecar de decir algo que a lo mejor no estoy cien por ciento consciente de ello; pero incluso la información que tiene INEGI, no es posible actualmente considerarla como de datos abiertos; porque a la hora de querer cruzarla, no te da. No te da porque, nos pasó hace poco, quisimos hacer una base de datos con el tema del sector femenino, en las mujeres, con relación a los empleos y con relación a la educación si tu cumples esas bases de datos no es el mismo, que las bases de datos son diferentes y que los datos no están estructurados de tal forma que te permitan hacer ese

cruce de información de manera tan fácil o tan rápida, por donde se le vea, no hay manera, yo digo que no hay datos abiertos aquí en este país.”

7. ¿Que herramientas utilizas para procesar las grandes cantidades de información en cada uno de los procesos?

“Todavía nos hace falta llegar a ese nivel, de tener que estar usando grandes cantidades de información; precisamente porque, una; porque no hay manera, no nos da la vida, no nos da cómo el tiempo; porque no tenemos quizás la experiencia, porque no somos expertos.

Por ejemplo, es usar STATA, o poder usar otro tipo de herramienta. SPSS, es, por ejemplo, un software que te permite procesar grandes cantidades de datos, lo hemos hecho en algunos trabajos, por ejemplo; en el proyecto de las carreras que se hizo, en proyectos como, en la misma UNAM que son bases de datos de más de doscientos mil datos, siento que todavía nos hace falta llegar a ese nivel.

Y cuando lo hemos hecho, en realidad en mi equipo no hay nadie que tenga esa experiencia para utilizar ese tipo de procesadores de bases de datos, y creo yo que tendríamos la capacidad de hacerlo si contráramos con un economista o con una analista de datos que fuera muy ducho, ósea no se tardan nada y al mismo tiempo porque las bases de datos con las hemos trabajado, no tiene como digamos esa complejidad, no son grandes bases de datos, pero bueno

tenemos el conocimiento de que se puede usar este tipo de software que permitirá usar esta gran cantidad de datos.”

7.1 ¿Otras herramientas que utilices?

“Pues el día a día es el Excel, utilizamos, por ejemplo; pues a veces hacemos cosas con Tableau pero no es muy común que, una, El Universal tiene un tema con las herramientas gratuitas, nosotros encantadas de usar Tableau, -inaudible-, Data Raw, Fusion Charts, qué otras herramientas que conocemos, para visualizar datos...”

...Hay muchas, pero ellos no quieren que se vayan los clics a otros sitios. Entonces, tú, embebes algo inmediatamente, cuando la gente le da clic, el clic se va a ese otro link, no al de nosotros; y eso implica que precisamente muchas de las herramientas que hacen se tienen que hacer en casa; con herramientas que tenemos dentro de la editorial; sino prácticamente es muy difícil que nos dejen hacer algo fuera. Porque, por ejemplo, hemos hecho cosas en d3, pero porque los archivos están dentro del servidor del periódico; o hemos hecho cosas en Carto, en Carto DB, pero aun así hay mucha resistencia a utilizar este tipo de herramientas. Además, como Carto lo acaba de demostrar, van cambiando; y a la hora de que van cambiando y evolucionando en el sitio del periódico, esos cambios se ven, o sea, y se ven medio feo y raro. Ya casi esa visualización si ya paso mucho tiempo, o simplemente se cae el servidor que es lo que nos ha pasado, o se desaparece la

herramienta que también nos ha pasado y eso es horrible, pero bueno son gajes del oficio.

Y pues nosotras, las otras herramientas que usamos bueno no sé si debería contar como herramientas, pero es algo que deberían usar todos los periodistas que usamos todos los días es hacer acceso a la información usar la plataforma Infomex, básicamente es eso.”

8. ¿Por ejemplo en Excel o cuando ya tienes la información que criterios utilizas para categorizar la información?

“Bueno, eso es como muy amplio, - ¿no? - Puede ser, entrar y usar varias hojas de cálculo, pero en distintos archivos; pero nosotros acostumbramos trabajar en cada base de datos. Por ejemplo, la descargamos - ¿no? - y algo que siempre te recomiendan -ya- todas las personas que han trabajado en Excel, es decir, copias el archivo y luego trabajas en uno nuevo; y después, por ejemplo, para mandarle la base de datos al infografista, debes descartar mucha información y mandarle otro archivo diferente al infografista de esa base de datos, ya trabajada y analizada.

Básicamente ese es el proceso que hacemos cuando nosotros queremos compartir la base de datos de manera pública al lector que ve el multimedia, casi siempre compartimos la base de datos que ya está limpia; es muy raro que compartamos la original. Sí deberíamos hacer un ejercicio para que vean como estaba antes y el

después; pero bueno, en realidad, ya es más bien de hacer algo así muy de nosotros. Porque en realidad si cambian las bases de datos de cómo estaban antes y después, y pues no sé si era un poco lo que querías escuchar.”

9. ¿Que dispositivos electrónicos utilizas tu para realizar la investigación?

“Bueno, eso es algo muy importante; porque resulta que nos hemos dado cuenta que las funciones de Excel no están de manera similar cuando estás trabajando en una PC que, en una Mac; nosotros por lo regular trabajamos en PC no trabajamos en Mac; pero también los diseñadores trabajan en PC -lo cual es espantoso para algunos diseñadores, pero bueno ahí les tocó vivir-; por ejemplo, para hacer pruebas, para ver cómo se ve el multimedia; hacemos pruebas en varios dispositivos, nuestros lectores entran desde el teléfono celular o desde el iPad y vean todo lo reportado en la visualización, no siempre se logra desafortunadamente porque tiene que ver mucho con la experiencia del diseñador web de armar un proyecto responsivo, o simplemente se queda en una sola versión, y cuando se queda en una sola versión es cuando no se ve bien en dispositivos móviles y básicamente es eso, ósea tratamos de usar todas las herramientas a nuestro alcance para poder usar, para poder visualizar para encontrar otro tipo de ejemplos de cómo hacer nuestro propio trabajo.”

10. ¿Desde tu perspectiva y la experiencia que tienes cual es la importancia que le ves al ejercicio del periodismo de datos?

“Pues es lo que les decía. El servicio social o el servicio a la comunidad; yo creo que es algo que nosotros como periodistas siempre tenemos como objetivo - ¿no? -, el revelar algo que no quieren que sea dicho, que sea revelado. O sea, descubrir una historia en la que tú estás haciendo esa investigación; y la otra es el servicio a la comunidad, yo creo que el Periodismo de Datos puede servir para las dos cuestiones y esa es la importancia porque, a raíz de que puedes sistematizar grandes cantidades de información, puedes encontrar patrones, puedes revelar ciertos comportamientos, puedes descubrir hallazgos que nadie había visto.

Creo que, en los ejemplos de Periodismo de Datos, hay muchas historias de éxito que rebelan o que tiran mentiras, - ¿no? - y que tiras verdades que supuestamente eran ciertas no lo son, para nada. No sé si recuerdan la historia de que en Estados Unidos la gente se gastaba mucho en rosquillas y decían que no tenían dinero para armas o chalecos antibalas. Resultaba que -ahora- ellos, se gastaban más del presupuesto en donas que en las armas, era inverosímil pero cierto. A la hora de hacer todo el análisis de la información, -y lo mismo puede ocurrir aquí-, te estás dando cuenta que quieres saber de que todos los días va el presupuesto para cambiar la bombilla o la bombilla de luz pero todos los días

sabemos que a alguno funcionario público le pagamos hasta tres veces más de lo que se merece ganar, entonces creo que sirve para eso y para explotar mucho la información que en una nota diaria, o de día a día no te deja satisfecho con la información que estás dando a conocer. “

Eduard Martín Borregon

Fecha: 20/10/2017

Periodista de datos. Inició la sección de Periodismo de Datos de El Periódico de Catalunya, realizó el Master en Periodismo de Investigación, Datos y Visualización por la URJC con el apoyo de Google. Actualmente trabaja en México en la ONG PODER donde coordina el departamento de tecnologías para la transparencia, haciendo proyectos basados en datos.

1. ¿Cómo definiría a un periodista de datos?

“Haber, yo entiendo que un periodista de datos...”

...Estamos hablando de una súper especialización, es decir; de la misma manera que podemos hablar –yo creo que hay que hablar del periodista de datos de la misma manera que hablamos- del fotoperiodista, - ¿no? – es decir; es en base una técnica que dominas, una técnica con mucha profundidad. Entonces, te especializas a partir de esta técnica, no quiere decir que no puedas hablar de datos sin ser periodista de datos, - ¿no? -. De la misma manera que tenemos reporteros que tienen la capacidad de salir a reportear y además tener buenas fotos sin ser fotoperiodistas, pues digamos, el periodista de datos -y aquí me separo un poco de la pregunta-, el periodismo de datos lo que es, es una técnica de aplicar técnicas de análisis de datos para el periodismo y entonces, estamos los súper

especializados -como es mi caso-, en que nuestra principal vertiente es la propia técnica que no la temática; aunque también nos acabamos especializando por temáticas, obviamente, - ¿no? -, pero que al final una de nuestras especializaciones es la técnica como sería la de un fotoperiodista o la Philip Roth de un videógrafo o la de un periodista narrativo.”

2. Mencione al menos 3 características que distinguen a un periodista de datos de uno de investigación.

“Es curioso porque, el Periodismo de Datos se ha asociado mucho, es decir hay varias vertientes, varias escuelas del Periodismo de Datos y curiosamente cada vez veo que tiene más fuerza la mezcla de periodismo de datos y periodismo de investigación, que también es la que yo practico, pero se podría hacer trabajo de datos desde la infografía, - ¿no? -, o evolucionar la infografía hacia el Periodismo de Datos que son dos disciplinas muy próximas. Entonces creo que es una tendencia; una tendencia bastante marcada, sobre todo porque, aquí deberíamos agarrar un poco de historia, - ¿no? -, cuando se habla de los antecedentes del Periodismo de Datos como el periodismo de precisión de Phillip Mayer o el CAR, el Computer Asistent Reporting de que empezaron aplicar. Pues son periodistas de investigación, que empezaron a aplicar su trabajo en bases de datos que han trabajado siempre con bases de datos pero empezaron a automatizarlas y agarrarlas

como periodismo de investigación, - ¿no? -, y yo creo que es más por una tradición de la que venimos, - ¿no? -, y de donde se empieza a desarrollar o uno de los campos de desarrollo.

Pero yo no diría que el Periodismo de Datos debe estar siempre vinculado al periodismo de investigación, - ¿no? -, el otro día veía una aplicación muy divertida no sé si era de la BBC, creo que no era de la BBC, no sé si era de la BBC, creo que no era de la BBC por la temática, en que te agarraban y te decían, “Estos versos, son versos de poetas históricos o son canciones de reggaeton”, que lo podríamos considerar como periodismo de datos, no, podríamos considerar a las new apps como periodismo de datos, pero no tenía nada de investigación.”

2.1 ¿Qué características te definen como periodista de datos a diferencia de un periodista de investigación?

“Creo que es el punto de vista, es decir, vamos a imaginarnos un listado telefónico, el periodista de investigación clásica lo que busca es una pista para luego agarrar el listado telefónico, agarrar este listado telefónico y ver cuál es el teléfono de la persona, - ¿no? -, o buscar el punto específico, pero necesita está pista, que lo dirige hasta aquí. Mientras que, el periodista de datos no tiene por qué empezar, no tienen este inicio sobre esta pista y lo que agarra es, que agarra todo el estado telefónico y empieza a decir, a ver, “¿Cuáles son las direcciones comunes?, ¿Qué dirección tiene más de

un teléfono? Uy, está dirección tiene más de cincuenta teléfonos, uy que raro, voy a ver qué pasa aquí” - ¿no? -, simplificando mucho. Es este enfoque o punto de partida distinto o técnica distinta.

Simplemente tiene una técnica distinta, o a lo mejor me gustaría decir distinta, en este caso tiene una técnica complementaria, - ¿no? -, es decir, complementa su técnica con otra, no, podemos decir que, con toda probabilidad y en eso aquí luego hay las mil y una excepciones; los mil y un casos, - ¿no? -, con toda probabilidad un periodista de investigación clásica tendrá una agenda de contactos kilométrica, - ¿no? -, mientras que un periodista de datos tendrá un listado de bases de datos kilométrico, para ir buscando.”

3. Describa brevemente cada uno de los pasos que lleva a cabo al realizar una investigación basada en datos.

“Mira, justo donde trabajo ahora donde trabajamos con una organización un poquito más clásica que nosotros; entonces llegaron, nos propusieron un tema a dos organizaciones distintas para que empezáramos a trabajar y a ver que podríamos hacer con ello. Entonces empezamos a trabajar y llegó el momento de empezar a plantear el trabajo, entonces yo para empezar a plantear el trabajo digo, bueno, antes de empezar a hacer cualquier cosa lo que he de ver sobre el tema que vamos a trabajar, ver que bases de datos tengo disponibles, como están formateadas, como las puedo cruzar, que

puedo ver, para a partir de aquí empezar a encontrar cruces, empezar a encontrar patrones y luego empezar a reportear encima de eso, - ¿no? -; y luego me encuentro con la otra organización que me dice “No, no, no los datos ya sé dónde están a mi lo que me interesa es empezar a buscar fuentes, y empezar a buscar fuentes para conseguir, esto, esto y esto”, yo digo espera, a ver espera aún no sabemos si los datos, yo no sé si los datos llegaran hasta donde tú quieras llegar, no, es decir, tú ya estás presuponiendo que tus enfoques son estos, mi enfoque es, primero analizo la base de datos y luego veré hacia donde voy, pero sin este primer análisis yo no puedo definir nada, estos datos los que luego me definirán me enseñaran hacia donde voy.”

4. Para realizar sus investigaciones, ¿qué tipo de bases de datos utiliza? , ¿Por qué?

“Depende de la investigación, depende muchísimo de la investigación. Últimamente nos hemos especializado bastante en PODER en contrataciones y en follow de money, es decir, seguir el dinero y seguir el cómo se va moviendo el dinero y como se van moviendo las cosas a través para hacer las investigaciones y últimamente hemos usado muchas bases de datos relacionadas con estos, pero como un hecho circunstancial, no un hecho principal que digamos.”

4.1 ¿Esas bases de datos, ustedes las generan o son públicas?

“Nos hemos encontrado de todo, trabajamos

con bases de datos públicas que no abiertas que no están en forma de open data, eso hay que recalcar, luego a veces encuentras bases de datos privadas, tenemos acceso a bases de datos de pago. A partir de las cuales nos da acceso a cierta información de forma mucho más rápida depende mucho del caso.”

4.2 ¿Realizan peticiones de información?

“Sí, también y muchas veces en el acceso de información lo triste es que, muchas veces las peticiones de información que hacemos es más para conseguir porque en la página web, no tenemos un formato abierto y antes de hacer un Web Scrapping y perder el tiempo, invertir el tiempo en hacer Web Scrapping decimos vamos a probar si nos dan está base de datos por transparencia, - ¿no? -, y también en muchos casos acabamos, no haciendo la búsqueda de las bases de datos sino haciendo el reporte, haciéndolo a través de peticiones de información, - ¿no? -, cosa que nos llega donde me ha llevado la base de datos a través de ahí, lo complemento con solicitudes de acceso a la información.”

5. Del proceso que describió anteriormente, mencione 3 problemáticas o dificultades que se suscitan durante su desarrollo.

“Yo hablaría para ser más de la falta de estándares y la falta de homogenización de las bases de datos, es decir, en un lugar encuentro el nombre de una forma, por ejemplo:

en primer apellido coma segundo apellido coma nombre, en otro lugar encuentro nombre y apellido separados, o en otro lugar es primer apellido y después el segundo apellido separado; luego por ejemplo en las bases de datos de Compranet esto ya lo tenemos publicado desde hace más de un año, a veces los nombres de las empresas no coinciden SA de CV con puntos, a veces los ponen sin puntos, después ponen una coma, a veces no ponen coma, eso te mete muchísimo trabajo en sistematización de la información, es decir en sistematizar y ordenar la información.

Un segundo gran problema es el tema es el acceso a la información, demasiadas veces el acceso a la información en México es, es vía, te devuelven PDFs generados, esto a que te devuelvan PDFs generados por lo que he hablo con varios funcionarios no es tanto un problema de mala fe o de querer molestar, sino que en realidad no tienen acceso al documento digital es un problema de archivo. El funcionario cuando agarra el archivo lo que haces es, agarra el documento, es decir termina un documento le da imprimir va a el archivo muerto deja en el archivo muerto el papel y entonces la unidad de transparencia la única manera que tiene de acceder a esto, es ir al archivo sacar el papel, escanearlo y mandarlo, es decir; no hay la versión digital.

Y un tercer problema es directamente ya la falta de información, está información no se puede encontrar, no es pública, hay que pagar por ella, por es decir, por ejemplo

es indignante nosotros tuvimos que pagar casi diez mil pesos para conseguir un contrato del nuevo aeropuerto, para eso pues tuvimos que pagar para tener la copia del contrato público, diez mil pesos para la impresión, lo cual es una locura, nosotros tenemos la suerte que nos podemos permitir eso como organización pero si no es que no te puedes permitir estos lujos.”

6. ¿Qué herramientas utiliza para procesar grandes cantidades de información?

“Por distintas circunstancias he tenido la suerte de casi siempre trabajar con programadores, lo cual me ha permitido, no tener que depender tanto de herramientas ya prefabricadas, en su momento se programar algunas cositas, aunque soy un programador muy malo, soy un reprogramador, solo malo, entonces si a veces he agarrado un código y lo he reprogramado yo uso mucho Excel, como te decía al principio, para mi es una herramienta muy natural, muy lógica de usar.”

6.1 ¿Qué herramienta consideras que se adapta más a tus necesidades?

“Para mi Excel y ahora uso por ejemplo en la oficina trabajamos con Linux, uso Open Calc y seamos honestos Excel es muy superior a Open Calc creo que es Libre Calc es muy superior Excel a Libre Calc en muchísimas cosas, es decir en capacidad de apertura de registros, en facilidad de uso de tablas dinámicas, incluso he llegado, para

que entiendas he llegado a programar pequeños algoritmos en Excel hay cosas que programo yo el algoritmo, la lógica del algoritmo con Excel testeo que funciona y lo paso al programador para que lo codifique en JavaScript. Ahora trabajamos en JavaScript pero he llegado a este proceso, -¿no? -, o a definir o decir muchas veces, participo en la definición de procesos lógicos de los algoritmos.”

7. ¿Qué características poseen las herramientas que mencionó anteriormente que hicieron que las seleccionara sobre otras?

“Excel tiene una interfaz gráfica, muy buena, muy intuitiva, JavaScript, en realidad es porque el programador quiso trabajar con JavaScript pero es el lenguaje del diablo. Entonces nunca yo hubiera elegido, sí que he hecho mis pinitos con Python o con R pero como digo la interfaz gráfica de no ver la tabla es una abstracción muy importante no tan fácil, por lo que he visto con R Estudio parece que se podría saltar o que podría dar un salto natural, saltar de Excel a R Estudio no, pero he programado. He hecho cursitos de R y tal y pero al final nunca he tenido la necesidad de hacer un trabajo en profundidad con R, lo cual no me ha ayudado nunca, es decir al final la programación la aprendes cada día, cada día, cada día, tocar cada día, tocar y como no lo toco cada día, pues no he acabado de aquí y luego vienen herramientas, toda la suite de herramientas de visualización pues la he usado todas,

desde los gráficos de Excel, Tableau, me supo muy mal que bajarán DataWrapper porque realmente era muy muy intuitivo, no me gusta Infogram porque me parece que infantiliza las cosas, pero también lo he usado en algún momento y la alternativa de Infogram creo que para un medio la alternativo de hacerte tu propio diseño puede ser una buena opción, no sé.”

8. ¿Qué criterios utiliza para seleccionar y categorizar grandes cantidades de información?

“Busco estándares ya preestablecidos por la comunidad o me baso ya en estándares ya preestablecidos, estándares de datos ya por suerte casi siempre hay un estándar en el cual basarse, - ¿no? -, o sobre todo en los casos más frecuentes hay un estándar a partir del cual basarse y en eso sí que fue un aprendizaje muy importante que tuvimos en ¿Quién es quién Wiki? que fue primero empezar a agarrar y construirnos nuestro propio esquema de datos, pero en lo que cruzas tres bases de datos juntas y no encajan, hay que empezar a decir, y ahora por que hice esto, pues es mucho mejor sentarse en las espaldas del gigante, - ¿no? -, y decir pues mira la comunidad lleva hasta aquí desarrollando un estándar de contrataciones, un estándar de beneficios reales, un estándar de propiedad, de relaciones tal, pues simplemente ya lo pensaron, ya les está funcionando pues usemos esto.”

8.1 ¿Utilizan alguna herramienta de obtención?

“Sí, mi proceso lógico es, obtención, limpieza, análisis y visualización o texto, dependiendo del objetivo que quiera tener, no, y digo esto porque, y en eso me incluyo, hubo un momento en el que parecía, un momento seguro te hablo sobre el 2014 que parecía que todo lo que teníamos que hacer en periodismo de datos era herramientas, no, es decir, te descargabas un base, la metías en Tableau le metías los cincuenta mil filtros para que el lector pudiera cruzar los datos en tropecientos maneras y ponías los datos de no sé qué, y al final creo que como estamos hablando de disciplina muy nueva, estamos hablando de la experimentación a partir de eso, creo que la disciplina, creo que es algo que debemos abandonar, - ¿no? -, porque en realidad el lector quiere una historia y en realidad lo que quiere es que le digas en donde está la historia, y si es necesario un gráfico, si complementa lo que estás diciendo, hagámoslo, pero si el gráfico no lo complementa, no lo hagamos, - ¿no? -, y a no ser que vayamos con un lector súper especializado, a un espacio muy especializado la herramienta es innecesaria, solo genera ruido y no acaba generando engagement con el lector, sino que solo lo ve, dice, ok que chido y se va.”


Anexo C: Storyboard de la sección de Obtención


Storyboard





Tema: Obtención


 Guion
 Animación




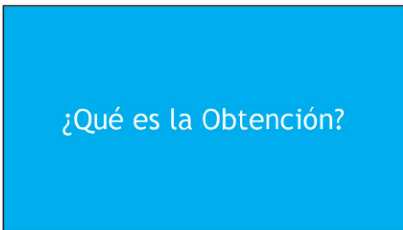
 Intro _____


 Se visualiza el intro de Dataista _____





 Para incursionar en la metodología propuesta por el Periodismo de Datos, es necesario conocer cada una de sus fases _____


 Icono de Obtención se animan los cuadros entran y salen de la caja. _____




 Empecemos por la obtención. _____

 La pregunta ¿Qué es la Obtención? entra poco a poco a cuadro. _____



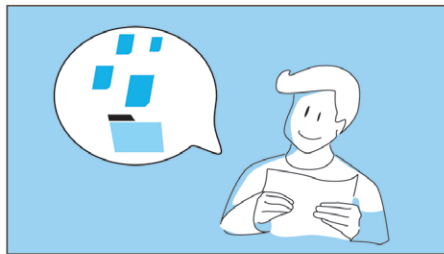
 Es la primera fase de la metodología. _____
 Es el momento en el que el periodista se plantea las fuentes de información de donde obtendrá las bases de datos que sustentarán su investigación. _____

 En la nube aparecen dos carpetas con signo de interrogación que desaparecen. _____

¿Qué se debe tomar en cuenta?

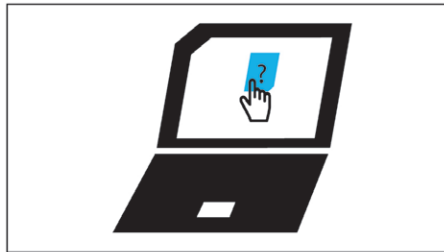
🎤 Durante la obtención, el periodista debe tomar en cuenta que:

▶ La pregunta ¿Qué se debe tomar en cuenta? entra a cuadro con un efecto.



🎤 Familiarizarse con el tipo de datos que existen respecto al tema que está investigando es importante. De esto dependerá la forma en la que obtendrá la información que requiere.

▶ Del folder desaparece el signo de interrogación y salen datos.



🎤 Como en toda fuente de información, es importante indagar sobre la metodología de recopilación de la base de datos que pretende explorar.

▶ El dato aparece en la computadora y la mano da click y desaparece el signo.



🎤 Saber si la fuente es confiable, si la información está actualizada y, sobre todo, el tipo de información que contiene te permitirá definir si la base de datos que obtuviste, es válida para su investigación.

▶ Da click aparece la paloma, en cada click la paloma, verificando.



🎤 La forma más común de obtener datos, es a través de solicitudes al gobierno o bien, descargarla de sitios web.

▶ Aparece una solicitud, después el logo de web y la descarga de archivo.



🎤 En el caso de México, el gobierno ha puesto a disposición de los usuarios el sitio datos.gob.mx. En él se albergan las bases de datos que las diferentes dependencias gubernamentales han puesto a disposición de los mexicanos.

▶ Zoom al sitio datos.gob.mx

Datos Abiertos

🎤 A las bases de datos dispuestas por los gobiernos para la consulta de sus ciudadanos, se les conoce como Datos Abiertos.

▶ Aparece cortinilla de Datos Abiertos



🎤 Se puede definir a los datos abiertos como aquellos datos que pueden ser utilizados y redistribuidos libremente por cualquier persona.

▶ Dato con candado abierto.



🎤 y que se encuentran sujetos, cuando más al requerimiento de atribución y de comportarse de la misma manera en que aparecen.

▶ Aparecen más datos abiertos con candados que están cerrados y abren.



🎤 Otra forma de obtener datos, es a través de búsquedas en Google.

▶ Página de búsqueda avanzada de Google, se hace zoom a búsqueda avanzada.



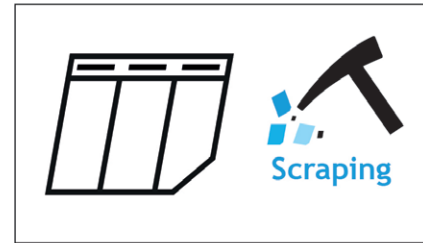
🎤 Un tip para que el buscador te dé solo bases de datos es, escribir la búsqueda y agregar al final ".xls" o ".csv". Así, Google buscará solo las bases de datos que tengan las características de tu búsqueda.

▶ Pantalla con Google Forms abierto, y se hace un close up a la aplicación.



🎤 En caso de que el gobierno no posea bases de datos sobre el tema de tu investigación, será necesario realizar peticiones vía transparencia para obtener la mayor documentación posible.

▶ Aparece el mapa de México al hacer click en la carpeta, aparece un fache y la libreta.



🎤 En estos casos, deberás armar la base de datos desde 0. Para facilitar el proceso, existen técnicas como el scraping. ¿Quieres saber de qué trata? Da clic al siguiente video.

▶ Base de datos limpia y aparece el logo de scraping con efecto.



🎤 Créditos

▶ Colectivo C datos Madic UAM Cuajimalpa

Anexo D: Código Fuente Herramienta Obtención

```

#!/usr/bin/env python3
# -*- coding: utf-8 -*-
#-----#
#Importacion de Librerias
#-----#
from flask import Flask,render_template,request, send_file, redirect, url_for, send_from_directory
from werkzeug.wrappers import Request, Response
from werkzeug import secure_filename
import PyPDF2
from urllib.request import urlopen
from urllib.error import HTTPError
from urllib.parse import urlparse
from bs4 import BeautifulSoup
from werkzeug import secure_filename
import requests
import os
import pandas as pd
from pandas.io.json import json_normalize
import json
from pandas import ExcelWriter
from pandas import ExcelFile
#-----#
#####
#-----#
#-----#
#Abre una instancia de Flask
#-----#
path=os.path.abspath("C:/Users/Erick/Desktop/obten/app/upload")
UPLOAD_FOLDER =path
ALLOWED_EXTENSIONS = set(['pdf'])
app = Flask(__name__)
app.config['UPLOAD_FOLDER'] = UPLOAD_FOLDER
# -----#
# Gestion de Errores
# -----#
@app.errorhandler(404)
def page_not_found(error):
    return 'Lo sentimos esta página no existe', 404
# -----#
# Ruta hacia la interfaz de la aplicación
# -----#
@app.route('/',methods=['GET'])
def formulario():
    return render_template('interfaz_obtencion.html')
# -----#
# Ruta 0: Obtencion de URL
# -----#

```

```

@app.route('/url',methods = ['GET', 'POST'])
def url():
    entrada=request.form['url']
    archivo=open('url.txt', 'w')
    for elemento in entrada:
        archivo.write(elemento)
    archivo.close()
    return render_template('url.html')
# -----
# Ruta 0.1: Obtencion de JSON
# -----
@app.route('/json',methods = ['GET', 'POST'])
def json():
    entrada=request.form['json']
    archivo=open('json.txt', 'w')
    for elemento in entrada:
        archivo.write(elemento)
    archivo.close()
    return render_template('json.html')
# -----
# Ruta 1: Obtencion de Parrafos
# -----
@app.route('/parrafo',methods = ['GET', 'POST'])
def parrafo():
    opcion=request.form['parrafo']
    liga=open('url.txt').read()

```

```

url=liga
html=urlopen(url)
bsObj=BeautifulSoup(html.read(), "lxml")
archivo=open('C:/Users/Erick/Desktop/obten/app/parrafo.txt', 'w')
for elemento in bsObj.find_all('p'):
    print (elemento)
    linea=str(elemento)
    archivo.write(linea)
    archivo.write('\n')
return render_template('salida_parrafo.html', texto=linea)
archivo.close()
# -----
# Descarga Parrafos
# -----
@app.route("/descarga_parrafo/", methods = ['GET'])
def descarga():
    if request.method == 'GET':
        tema=request.form['descarga_parrafo']
        return send_from_directory(directory='C:/Users/Erick/Desktop/obten/app/parrafo.txt', filename='parrafo.txt',as_attachment=True)
@app.route('/descargar/', methods = ['GET', 'POST'])
def descargar():
    path=os.path.abspath('C:\\Users\\Erick\\Desktop\\obten\\app\\parrafo.txt')
    return send_file(path,attachment_filename='parrafo.txt')
# -----

```

```

# Ruta 2: Obtencion de Tablas
# -----
@app.route('/tablas',methods = ['GET', 'POST'])
def tablas():
    request.form['tablas']
    liga=open('url.txt').read()
    url=liga
    html=urlopen(url)
    bsObj=BeautifulSoup(html.read(), "lxml")
    archivo=open('tablas.txt', 'w')
    for elemento in bsObj.find_all('table'):
        print (elemento)
        linea=str(elemento)
        archivo.write(linea)
        archivo.write('\n')
    return render_template('salida_tablas.html', title=linea)
    archivo.close()

# -----
# Descarga Tablas
# -----
@app.route("/descargar_tablas", methods = ['GET'])
def descargar_tablas():
    if request.method == 'GET':
        tema=request.form['descarga_tablas']
        return send_from_directory(directory='C:/Users/Erick/Desktop/obten/app/tablas.txt', filename='tablas.txt',as_attachment=True)

```

```

@app.route('/descarga2/', methods = ['GET', 'POST'])
def descarga2():
    path=os.path.abspath('C:\\Users\\Erick\\Desktop\\obten\\app\\tablas.txt')
    return send_file(path,attachment_filename='tablas.txt')

# -----
# Ruta 3: Obtencion de Enlaces
# -----
@app.route('/enlaces',methods = ['GET', 'POST'])
def enlaces():
    request.form['enlaces']
    liga=open('url.txt').read()
    url=liga
    html=urlopen(url)
    bsObj=BeautifulSoup(html.read(), "lxml")
    archivo=open('enlaces.txt', 'w')
    for elemento in bsObj.find_all('a'):
        print (elemento)
        linea=str(elemento)
        archivo.write(linea)
        archivo.write('\n')
    return render_template('salida_enlaces.html', contador=linea)
    archivo.close()

# -----
# Descarga Enlaces

```



```

# -----
@app.route("/descargar_enlaces", methods = ['GET'])
def descargar_enlaces():
    if request.method == 'GET':
        tema=request.form['descarga_enlaces']

        return send_from_directory(directory='C:/Users/Erick/Desktop/obten/app/enlaces.txt', filename='enlaces.txt',as_attachment=True)

@app.route('/descarga3/', methods = ['GET', 'POST'])
def descarga3():
    path=os.path.abspath('C:\\Users\\Erick\\Desktop\\obten\\app\\enlaces.txt')

    return send_file(path,attachment_filename='enlaces.txt')
# -----
# Ruta 4: Obtencion de Encabezados
# -----
@app.route('/encabezado',methods = ['GET', 'POST'])
def encabezado():
    url=request.form['encabezado']
    liga=open('url.txt').read()
    url=liga
    html=urlopen(url)
    bsObj=BeautifulSoup(html.read(), "lxml")
    archivo=open('encabezado.txt', 'w')
    for elemento in bsObj.find_all('h1'):
        print (elemento)
        linea=str(elemento)
        archivo.write(linea)

```

```

        archivo.write('\n')

        return render_template('salida_encabezados.html', texto=linea)

        archivo.close()
# -----
# Descarga Encabezados
# -----
@app.route("/descargar_encabezados", methods = ['GET'])
def descargar_encabezados():
    if request.method == 'GET':
        tema=request.form['descarga_encabezados']

        return send_from_directory(directory='C:/Users/Erick/Desktop/obten/app/encabezado.txt', filename='encabezado.txt',as_attachment=True)

@app.route('/descarga4/', methods = ['GET', 'POST'])
def descarga4():
    path=os.path.abspath('C:\\Users\\Erick\\Desktop\\obten\\app\\encabezado.txt')

    return send_file(path,attachment_filename='encabezado.txt')
# -----
# Ruta 5: EXTRACCION DE TEXTO DE UN ARCHIVO PDF
# -----
# Ruta 5.1: Cargar Archivo
# -----
@app.route("/subir", methods = ['GET', 'POST'])
def subirArchivo():
    nombref= request.files['archivo']

```

```

filename = secure_filename(nombref.filename)

filename='archivo.pdf'

nombref.save(os.path.join(app.config['UPLOAD_FOLDER'],
filename))

return render_template('descarga.html')

@app.route('/extraer', methods = ['GET', 'POST'])
def extraer():

nombref= request.form['archivo']

pdf_file = open('C:/Users/Erick/Desktop/obten/app/up-
load/archivo.pdf', 'rb')

read_pdf = PyPDF2.PdfFileReader(pdf_file)

number_of_pages = read_pdf.getNumPages()

page = read_pdf.getPage(0)

page_content = page.extractText()

page_content=page_content.encode('utf-8')

return render_template('salida_pdf.html', texto=page_
content)

# -----
# Ruta 6: Extracción de un Archivo JSON
# -----

@app.route('/JSON', methods = ['GET', 'POST'])
def JSON():

nombref= request.form['JSON']

liga=open('url.txt').read()

url=liga

try:

e=pd.read_json(requests.get(url).text, ori-

```

```

ent='columns')

archivo=e.to_excel('salida_json.xlsx')

html=e.to_html()

return render_template("salida_json_1.html",
salida=html)

except:

url=liga

d = json.loads(requests.get(url).text)

df = pd.DataFrame.from_dict(json_normalize(d),
orient='columns')

archivo2=df.to_excel('salida_json.xlsx')

html=df.to_html()

return render_template("salida_json_2.html", sa-
lida=html)

archivo.close()

# -----
# Descarga JSON
# -----

@app.route("/descargar_json", methods = ['GET'])
def descargar_json():

if request.method == 'GET':

tema=request.form['descargar_json']

return send_from_directory(directory='C:/Users/
Erick/Desktop/obten/app/salida_json.xlsx', filename='salida_
json.xlsx',as_attachment=True)

@app.route('/descarga5/', methods = ['GET', 'POST'])

```

```

def descarga5():
    if request.method == 'POST':
        path=os.path.abspath('C:\\Users\\Erick\\Desk-
top\\obten\\app\\salida_json.xlsx')

        return send_file(path,attachment_filename='sali-
da_json.xlsx')

# -----
# Correr Aplicacion
# -----

if __name__ == '__main__':
    app.run() #!/usr/bin/env python3

```

Anexo E: Código Fuente Herramienta Limpieza

```

# Coding= utf-8

#-----#

#Importación de Librerias
#-----#

from flask import Flask,render_template,request, send_file, redi-
rect, url_for, send_from_directory

from werkzeug.wrappers import Request, Response

from werkzeug import secure_filename

import pandas as pd

from pandas import ExcelWriter

from pandas import ExcelFile

import os

#-----#

#Abre una instancia de Flask

#-----#

path=os.path.abspath("C:/Users/Erick/Desktop/Analizal/app/up-
load")

UPLOAD_FOLDER =path

ALLOWED_EXTENSIONS = set(['xlsx'])

app = Flask(__name__)

app.config['UPLOAD_FOLDER'] = UPLOAD_FOLDER

# -----
# Gestion de Errores
# -----

@app.errorhandler(404)

```

```

def page_not_found(error):
    return 'Lo sentimos esta página no existe', 404
# -----
# Ruta hacia la interfaz de la aplicación
# -----
@app.route("/", methods = ['GET', 'POST'])
def home():
    return render_template("Interfaz_Limpieza.html")
# -----
# Ruta 1: Cargar Archivo
# -----
@app.route("/subir", methods = ['GET', 'POST'])
def subirArchivo():
    nombref= request.files['archivo']
    filename = secure_filename(nombref.filename)
    nombref.save(os.path.join(app.config['UPLOAD_FOLDER'],
'archivo.xlsx'))
    return render_template('salida.html')
# -----
# Ruta 2: Limpieza de los datos
# -----
@app.route("/tipoDatos", methods = ['GET', 'POST'])
def tipoDatos():
    tema=request.form['tipoDatos']
    archivo=pd.ExcelFile('C://Users//Erick//Desktop//Anali-
zal//app//upload//archivo.xlsx', sheetname=0)
    df=archivo.parse(0)[0:10]
    html=df.to_html(classes='estilo')
    hojas=archivo.sheet_names
    columnas=df.columns.tolist()
    return render_template('Resultados_Limpieza.html', sa-
lida1= html, salida2=hojas, salida8=columnas)
@app.route("/tipoDatos2", methods = ['GET', 'POST'])
def tipoDatos2():
    hoja=request.form['hoja']
    campo=request.form['campo']
    caracter=request.form['caracter']
    sustituto=request.form['sustituto']
    archivo=pd.ExcelFile('C://Users//Erick//Desktop//Anali-
zal//app//upload//archivo.xlsx', sheetname=0)
    df=archivo.parse(hoja)
    df[campo] = df[campo].replace([caracter], sustituto)
    writer = pd.ExcelWriter('C:\\Users\\Erick\\Desktop\\
Limpiar\\app\\limpio.xlsx', engine='xlsxwriter')
    df.to_excel(writer, sheet_name='Hoja1')
    writer.save()
    hojas=archivo.sheet_names
    columnas=df.columns.tolist()
    return render_template('Resultados_Limpieza.html', sa-
lida2=hojas, salida8=columnas)
# -----
# Descarga
# -----
@app.route("/descarga", methods = ['GET'])

```

```

def descarga():
    if request.method == 'GET':
        tema=request.form['descarga']

        return send_from_directory(directory='C:\\Users\\Erick\\Desktop\\Limpiar\\app\\', filename='limpio.xlsx', as_attachment=True)

@app.route('/descargar/', methods = ['GET', 'POST'])
def descargar():

    path=os.path.abspath('C:\\Users\\Erick\\Desktop\\Limpiar\\app\\limpio.xlsx')

    return send_file(path,attachment_filename='limpio.xlsx')

# -----#

# Correr Aplicación

# -----#

if __name__ == "__main__":

    app.run(debug=True)

#!/usr/bin/env python3

# Coding= utf-8

# -----#

#Importación de Librerías

# -----#

from flask import Flask,render_template,request, send_file, redirect, url_for, send_from_directory

from werkzeug.wrappers import Request, Response

from werkzeug import import secure_filename

import pandas as pd

from pandas import ExcelWriter

```

Anexo F: Código Fuente Herramienta Análisis

```

from pandas import ExcelFile

import os

# -----#

#Abre una instancia de Flask

# -----#

path=os.path.abspath("C:/Users/Erick/Desktop/Analizar/app/upload")

UPLOAD_FOLDER =path

ALLOWED_EXTENSIONS = set(['xlsx'])

app = Flask(__name__)

app.config['UPLOAD_FOLDER'] = UPLOAD_FOLDER

# -----#

# Gestion de Errores

# -----#

@app.errorhandler(404)

def page_not_found(error):

    return 'Lo sentimos esta página no existe', 404

# -----#

# Ruta hacia la interfaz de la aplicación

# -----#

@app.route("/", methods = ['GET', 'POST'])

def home():

    return render_template("Interfaz_Analizar.html")

```

```

# -----
# Ruta 1: Cargar Archivo
# -----
@app.route("/subir", methods = ['GET', 'POST'])
def subirArchivo():
    nombref= request.files['archivo']
    filename = secure_filename(nombref.filename)
    nombref.save(os.path.join(app.config['UPLOAD_FOLDER'],
'archivo.xlsx'))

    return render_template('salida.html')
# -----
# Ruta 2: Impresión de los datos
# -----
@app.route("/tipoDatos", methods = ['GET', 'POST'])
def tipoDatos():
    tema=request.form['tipoDatos']
    archivo=pd.ExcelFile('C://Users//Erick//Desktop//Anali-
zal//app//upload//archivo.xlsx', sheetname=0)

    df=archivo.parse(0)
    hojas=archivo.sheet_names
    columnas=df.columns.tolist()

    tipo=df.shape
    info=df.dtypes
    des=df.describe()
    tipo2=type(des)
    archivo=des.to_excel('archivo.xlsx')
    html=des.to_html(classes='estilo')

```

```

        return render_template('Resultados_Analisis.html',sali-
dal=html, salida2=columnas, salida6=tipo, salida7=info, sali-
da8=hojas)

        df.close()

@app.route("/tipoDatos2", methods = ['GET', 'POST'])
def tipoDatos2():
    tema=request.form['opcion']
    archivo=pd.ExcelFile('C://Users//Erick//Desktop//Anali-
zal//app//upload//archivo.xls', sheetname='Sheet1')

    df=archivo.parse(tema)
    hojas=archivo.sheet_names
    columnas=df.columns.tolist()

    tipo=df.shape
    info=df.dtypes
    des=df.describe()
    tipo2=type(des)
    archivo=des.to_excel('archivo.xlsx')
    html=des.to_html(classes='estilo')

    return render_template('Resultados_Analisis.html',sali-
dal=html, salida2=columnas, salida6=tipo, salida7=info, sali-
da8=hojas)

    df.close()

# -----
# Descarga
# -----
@app.route("/descarga", methods = ['GET'])
def descarga():
    if request.method == 'GET':

```

```

        tema=request.form['descarga']

        return send_from_directory(directory=app.config['UPLOAD_FOLDER'], filename='archivo.xlsx', as_attachment=True)

@app.route('/descargar/', methods = ['GET', 'POST'])

def descargar():

    path=os.path.abspath('C:\\Users\\Erick\\Desktop\\Anализa\\app\\archivo.xlsx')

    return send_file(path,attachment_filename='archivo.xlsx')

# -----
# Correr Aplicación
# -----

if __name__ == "__main__":

    app.run(debug=True)

from flask import Flask,render_template,request

from werkzeug import secure_filename

import pygal

from pygal.style import Style

from pygal.style import DefaultStyle

from pygal.maps.world import COUNTRIES

import json

app = Flask(__name__)

custom_style = Style(

    colors=( '#991515', '#1cbc7c' ),

    background='#d2ddd9' )

# -----

# Default route for the charting application

```

Anexo G: Código Fuente Herramienta Visualización

```

# -----

@app.route("/")

def home():

    return render_template("Interfaz_Vis.html")

# -----
# Charting route which displays the bar chart
# -----

@app.route("/bar", methods = ['GET', 'POST'])

def bar():

    tema=request.form['bar']

    with open('bar.json','r') as bar_file:

        data = json.load(bar_file)

        chart = pygal.Bar(style = custom_style)

        mark_list = [x['mark'] for x in data]

        tourn_list = [x['tournament'] for x in data]

        chart.add('Year',mark_list)

        chart.add('Marcador',tourn_list)

        chart.x_labels = [x['year'] for x in data]

        chart.render_to_file('static/images/bar_chart.svg')

        img_url = 'static/images/bar_chart.svg?cache='

        return render_template('linea.html',image_url = img_url)

@app.route("/line", methods = ['GET', 'POST'])

```

```

def line():tema=request.form['line']

    chart = pygal.Bar(print_values=True, style=DefaultStyle(value_font_family='googlefont:Raleway',value_font_size=30,value_colors=('white',)))

    chart.add('line', [0, 12, 31, 8, -28, 0])

    chart.render_to_file('static/images/linea.svg')

    img_url = 'static/images/linea.svg?cache='

    return render_template('linea.html',image_url = img_url)

@app.route("/hist", methods = ['GET', 'POST'])
def hist():

    tema=request.form['hist']

    hist = pygal.Histogram()

    hist.add('Wide bars', [(5, 0, 10), (4, 5, 13), (2, 0, 15)])

    hist.add('Narrow bars', [(10, 1, 2), (12, 4, 4.5), (8, 11, 13)])

    hist.render_to_file('static/images/hist.svg')

    img_url = 'static/images/hist.svg?cache='

    return render_template('linea.html',image_url = img_url)

@app.route("/map", methods = ['GET', 'POST'])
def map():

    tema=request.form['map']

    worldmap_chart = pygal.maps.world.World()

    worldmap_chart.title = 'Some countries'

    worldmap_chart.add('F España', ['fr', 'fi'])

    worldmap_chart.add('M Mexico', ['ma', 'mc', 'md', 'me', 'mg', 'mk', 'ml', 'mm', 'mn', 'mo', 'mr', 'mt', 'mu', 'mv', 'mw', 'mx', 'my', 'mz'])

    worldmap_chart.add('U Inglaterra', ['ua', 'ug', 'us', 'uy', 'uz'])

    worldmap_chart.render_to_file('static/images/map.svg')

    img_url = 'static/images/map.svg?cache='

    return render_template('linea.html',image_url = img_url)

if __name__=="__main__":

    app.run()

```