datalab

POLÍTICAS PÚBLICAS
BASADAS EN EVIDENCIA

SISTEMATIZACIÓN, ANÁLISIS Y VISUALIZACIÓN DE AGRESIONES CONTRA PERSONAS DEFENSORAS DE DERECHOS HUMANOS Y PERIODISTAS

datalab
POLÍTICAS PÚBLICAS
BASADAS EN EVIDENCIA

SEGOB
SECRETARÍA DE GOBERNACIÓN

SISTEMATIZACIÓN, ANÁLISIS Y VISUALIZACIÓN DE AGRESIONES CONTRA PERSONAS DEFENSORAS DE DERECHOS HUMANOS Y PERIODISTAS

UNIDAD DE PREVENCIÓN SEGUIMIENTO Y ANÁLISIS
COORDINACIÓN DE ESTRATEGIA DIGITAL NACIONAL
LABORATORIO NACIONAL DE POLÍTICAS PÚBLICAS

ABRIL 2017

MÉXICO
PRESIDENCIA DE LA REPÚBLICA

MÉXICO DIGITAL

CIDE

Laboratorio Nacional de Políticas Públicas

Índice

Introducción	04
Descripción del proyecto	05
Cronograma	06
Revisión de literatura	07
Desarrollo y evolución del proyecto	08
Metodología	12
Resultados	19
Conclusión	28
Siguientes pasos	29
Evaluación	31
Referencias	35
Anexos	36

Introducción

NECESIDADES DE LA DEPENDENCIA

De acuerdo con la Ley para la Protección de Personas Defensoras de Derechos Humanos, la Unidad de Prevención Seguimiento y Análisis es la dependencia encargada de realizar el monitoreo nacional de agresiones contra defensoras de derechos humanos y periodistas en ejercicio de sus labores (en adelante PDDHyP).

Con este fin, en el transcurso de esta administración se creó el Mecanismo de Protección para Personas Defensoras de Derechos Humanos y Periodistas. El Mecanismo tiene el objetivo de "implementar y operar las medidas de prevención, medidas de protección y medidas urgentes de protección que garanticen la vida, integridad, libertad y seguridad de las personas que se encuentren en situación de riesgo".

El Mecanismo no solo debe recopilar información acerca de agresiones en contra de PDDHyP, sino que también tiene que usar esta información para mejorar la atención y el cuidado de su población objetivo. Para esto, el Mecanismo debe recopilar y sistematizar información, identificar patrones de agresiones, así como elaborar mapas de riesgos para los grupos vulnerables que atiende. La idea detrás de esto es tratar de generar políticas públicas que protejan a las PDDHyP de forma más eficaz.

El cumplimiento de este encargo implica múltiples tareas: a) la detección periódica de las agresiones; b) la sistematización de dicha información; c) el análisis de las agresiones, tomando en cuenta el contexto político, económico y social que interviene en la generación de dicha violencia; d) la georreferenciación de las agresiones, entre otras. Adicionalmente, para el desarrollo de un análisis preciso de las agresiones contra PDDHyP resulta necesario recopilar información a lo largo de un periodo significativo de tiempo. Dada la gran cantidad de violencia contra estos grupos, este ejercicio resulta en un enorme cúmulo de información por recolectar, sistematizar y analizar. En consecuencia, el desafío radica en la generación de un programa capaz de detectar eventos de agresión con rapidez y eficiencia.

En principio, la Unidad de Prevención Seguimiento y Análisis comenzó el trabajo de monitoreo de fuentes periodísticas junto con la recopilación de información sobre eventos de agresión ocurridos durante 2015. No obstante, se carece de una metodología clara y objetiva, así como de datos más allá del año señalado.

MOTIVACIÓN DE SELECCIÓN DE PARTE DE DATALAB

Este proyecto se escogió por una varios distintos factores. Lo más importante es que se vio una necesidad del Mecanismo que podía ser atendida a través de la tecnología y el análisis de datos: generación de una base de datos de agresiones en contra de PDDHyP y la utilización de esta misma para mejorar la protección de las PDDHyP.

Actualmente el Mecanismo ha recopilado la información de agresiones en contra de PDDHyP de forma manual utilizando "Google Alerts" como herramienta principal. Este método es muy tardado y dada la falta de recurso económicos y humanos del Mecanismo, no se ha podido utilizar para generar una base de datos de agresiones a PDDHyP desde antes de 2015.

Datalab vio que un desarrollo tecnológico podría acelerar la recopilación de información y así completar una base de datos que fuese verdaderamente útil en términos de política pública.

Lo más importante, sin embargo, fue que el análisis de datos podría ser apalancado para generar política pública basada en evidencia. La base de agresiones contra de PDDHyP podría ser relacionada con otras fuentes de información y así generar insumos esenciales para la estrategia de prevención y protección del Mecanismo. Es decir, el análisis de datos podría generar evidencia para la generación de políticas públicas más eficaces. Este en realidad es el objetivo de Datalab.

Por último, el equipo de Datalab consideró que este proyecto representaba la oportunidad para trabajar en un tema de suma importancia para el país. Las violaciones a los derechos humanos y el acoso al periodismo impiden el acceso a la justicia, limitan la libertad de expresión y merman la confianza en las instituciones públicas. El poder apoyar al Mecanismo, por ende, representaba una gran oportunidad para contribuir a una de las problemáticas más importantes para el país.

Descripción del proyecto

OBJETIVO DEL PROYECTO

Objetivo general: Construir un modelo de análisis para la prevención de posibles violaciones a los derechos humanos de PDDHyP a partir de una base de datos que integre diferentes fuentes de información y sea georreferenciable para el desarrollo futuro de una herramienta de autoprotección para PDDHyP.

Objetivos específicos:

- Definición de conceptos básicos que guían el proyecto y su operacionalización subsecuente.
- Elaboración de una metodología objetiva para la recopilación de información sobre eventos de agresión contra PDDHyP.
- Generación de un software para la sistematización de información sobre eventos de agresión contra PDDHyP.
- Construcción de una base de datos de eventos de agresión contra PDDHyP durante el periodo 2009-2016.
- Análisis estadístico de los determinantes de los eventos de agresión contra PDDHyP durante el periodo 2009-2016.
- Visualización de los eventos de agresión contra PDDHyP durante el periodo 2009-2016.

PRODUCTOS ENTREGABLES

En un periodo de cuatro meses se generaron los siguientes productos, que serán a su vez entregados a la Unidad de Prevención Seguimiento y Análisis:

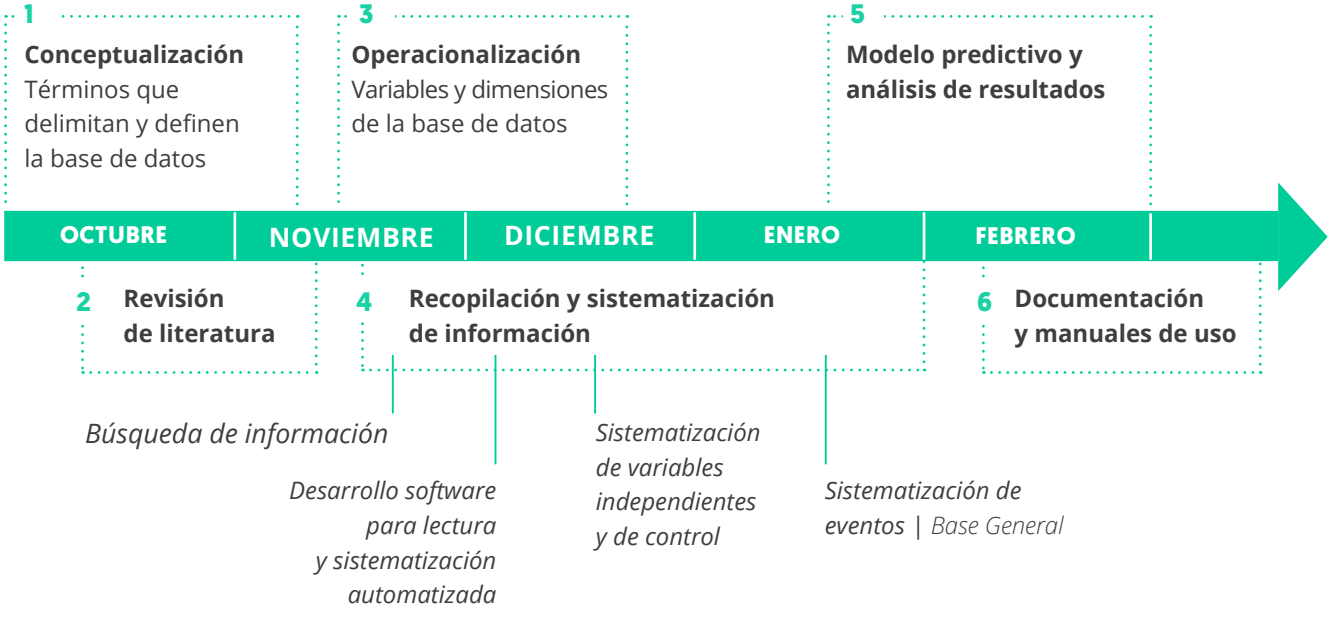
1. Manual de conceptualización, operacionalización y metodología detrás de la creación de la base de datos de eventos de agresión contra PDDHyP
   - Archivo en formato **.docx:"ManualDocumentacionBaseAgresiones PDDHyP"**
2. Algoritmos para la búsqueda y recolección de notas y reportajes en periódicos y archivos en línea de organizaciones de la sociedad civil
3. Archivos de texto que recopilan las notas periodísticas y alertas de organizaciones de atención a PDDHyP que alimentaron la base de

datos general sobre eventos de agresión

- Archivo en formato **.xls** que recopila *todas* las notas que identificó y capturó nuestro algoritmo: **"Notas2432"**
- Archivo en formato .xls que recopila las notas *depuradas* para la construcción de la base de datos: **"Notas1872"**

4. Diccionarios refinados de términos de lectura del software y códigos detrás de la generación de los mismos
   - Metodología de construcción de diccionarios
   - Código para uso de Word2vec
   - Código para generación de frecuencias de palabras
   - Código de N-gramas
   - Diccionario apilado (con toda la información de los otros diccionarios)
   - Diccionario de actores
   - Diccionario de estados y municipios
   - Diccionario de nombres propios
   - Diccionario de verbos
5. Software generado para la lectura y sistematización automatizada de eventos de agresión contra PDDHyP
6. Base de datos de eventos de agresión con información sobre sus características básicas: tipo de agresión, tipo de víctima, ubicación y fecha.
   - Archivo en formato **.xls: "agresiones_por_municipio.xls"**
7. Visualizador (*dashboard*)
   - Carpeta con archivos para la ejecución del visualizador: "Dashboard"
8. Modelos estadísticos sobre los factores de riesgo que explican las agresiones contra PDDHyP y su respectivo análisis
   - Ver sección de "Resultados" en este documento

Cronograma



Revisión de literatura

El estudio de las agresiones en contra de las personas defensoras de derechos humanos y periodistas debe partir de los trabajos realizados sobre la violencia ejercida tanto por el estado como por grupos privados (e.j., crimen organizado, corporaciones o empresas, específicamente en el caso de las PDDH relacionadas con el medio ambiente y recursos naturales). En términos generales, los estudios académicos se han enfocado en examinar la forma en que el contexto político, social y económico influye en la violencia dirigida en contra de nuestros dos grupos de interés. A continuación se describe la lógica de estos distintos argumentos.

Un primer grupo de la literatura académica se ha concentrado en entender la relación entre los tipos de regímenes políticos y la represión. Dentro de estos grupos, hay dos argumentos principales. Por un lado, hay investigaciones que concluyen que entre mayores sean los índices de democracia, la probabilidad de que un gobierno reprima a los individuos bajo su jurisdicción, disminuirá significativamente (Davenport y Amstrong, 2004; Bueno de Mesquita et al., 2005). La relación negativa entre democracia y represión se debe principalmente a que en un régimen democrático hay: a) una oposición política y una prensa libre capaces de denunciar abusos del gobierno (Whitten-Woodring y James, 2012), así como también, b) mecanismos de rendición de cuentas efectivos que reducen las injerencias en los derechos humanos (Bueno de Mesquita et al., 2005; Davenport, 2007). En contraste, un gobierno autoritario posee baja participación política y una oposición débil, además, generalmente, los mecanismos de rendición de cuentas son nulos y por lo tanto, las conductas represivas tienen un costo muy bajo.

Por otro lado, también existe el argumento de que la relación entre democracia y represión del estado no es lineal y que los regímenes más represivos son aquellos que muestran niveles intermedios de democracia (Fein, 1995; Regan y Henderson, 2002). Esto se debe a que las semi-democracias se encuentran en un punto intermedio en el que muchas veces las instituciones son débiles no existen los canales adecuados para canalizar las demandas de la sociedad y por último, el sistema de pesos y contrapesos no funciona correctamente. Así, al contar con instituciones y mecanismos débiles, los gobiernos en regímenes semi-democráticos son propensos a reprimir cuando un grupo disidente exige y demanda acciones por parte del Estado.

Un segundo grupo de trabajos argumenta que la represión política depende del nivel de amenazas que enfrentan los gobiernos no-democráticos (Davenport, 1995; Gartner y Regan, 1996; Regan y Henderson, 2002; Von Doeep y Young, 2013). Principalmente, la acción colectiva de la sociedad civil en la forma de protestas, por ejemplo, puede resultar amenazante para regímenes no-democráticos, generando incentivos a la censura o agresión contra las personas que estimulen o provean información sobre dichas manifestaciones públicas (King et al., 2013). Si bien escribir o articular críticas contra el gobierno no convierte al defensor o al periodista automáticamente en blanco de una agresión, el contenido de dichas críticas podría fomentar o facilitar la acción colectiva y así, provocar una

respuesta violenta por parte del Estado. El gobierno busca de esta forma romper las redes que sustentan a los sectores críticos dentro del gremio periodístico o de derechos humanos, por ejemplo.

Un tercer grupo de estudios sostiene que el contexto en el que se desarrolla profesionalmente el periodista o defensor importa para explicar la represión contra ellos. Por ejemplo, un contexto de violencia criminal tiene la capacidad de aumentar el riesgo en la seguridad de los defensores y periodistas. En el caso de los periodistas, Holland y Ríos (2015) encuentran que la probabilidad de ocurrencia de un ataque contra periodistas aumenta cuando grupos rivales del crimen organizado habitan el mismo territorio. La principal explicación a este fenómeno es que las zonas donde sólo existe un grupo criminal dominante podrá controlar más fácil el flujo de información. Por el contrario, cuando existe una rivalidad entre bandas criminales, la capacidad para controlar la información se debilita y se dificulta (Holland y Ríos, 2015). En este sentido, el ataque contra la prensa está determinado, en parte, por la rivalidad de grupos criminales. Si bien no hay trabajos específicos que analicen el impacto directo de la violencia criminal en el sector de defensa de derechos humanos, podríamos pensar en una lógica similar para el caso de los activistas de dicho sector. Los grupos criminales podrían tener incentivos para acallar a aquellos defensores de derechos humanos que realicen denuncias públicas de sus delitos, diseminando información sobre sus actividades y llamando la atención de las autoridades, lo cual podría afectar la libertad de acción e impunidad con la que operan estas organizaciones.

Finalmente, un cuarto grupo de estudios revela que las características económicas también son relevantes en el fortalecimiento de los derechos humanos. La globalización económica está asociada con un mejor comportamiento en el ámbito de derechos humanos (Franklin, 2008; Dutta y Roy, 2009). Los gobiernos no quieren perder lazos comerciales ni la interdependencia económica que tienen con otros países y para ello deben demostrar estabilidad política y respeto a la legalidad. En este sentido, se ha encontrado que las economías más aisladas tienen mayor probabilidad de cometer agresiones contra sus ciudadanos (Harff, 2003). Además, los países con mayor dependencia económica del extranjero tienden a reprimir en menor magnitud (Franklin 2008). Asimismo, la inversión extranjera directa está positivamente relacionada con la protección a la libertad de prensa (Dutta y Roy, 2009).

Desarrollo y evolución del proyecto

El proceso de recolección y sistematización de datos sobre las agresiones contra PDDHyP se desarrolló a lo largo cuatro meses que incluyeron diversas etapas con distintos niveles de complejidad y retos que se resolvieron sobre la marcha. A continuación se describen estos procesos y la evolución del proyecto a lo largo de este periodo, así como algunos de los obstáculos que se presentaron y la manera en que se atendieron o resolvieron.

I. CONCEPTUALIZACIÓN Y OPERACIONALIZACIÓN DE LAS AGRESIONES CONTRA PDDHYP

Toda base de datos debe partir de una conceptualización clara del fenómeno que busca sistematizar. Sin conceptos claros, fácilmente se generan inconsistencias a lo largo del proceso de recopilación y captura de datos. Una conceptualización clara de las variables y categorías que incluye una base de datos nos permite blindar nuestro trabajo de críticas y lo fortalece enormemente. Más aún, para hacer sustentable nuestras bases de datos a lo largo del tiempo, nuestras definiciones y criterios tienen que ser lo suficientemente claros y transparentes para ser replicados por otros más adelante. En este proyecto fue esencial definir tres conceptos en particular: 1) el tipo de víctimas, 2) el tipo de violencia y c) el tipo de agresiones.

Tras definir estos conceptos resultaba esencial identificar las formas de medir cada una las sus características y dimensiones de las agresiones contra PDDHyP. Se trata, por tanto, del proceso de operacionalización, lo cual implicaba delimitar de manera más específica: 1) los tipos de derechos vulnerados relevantes para la base; 2) los actores perpetradores de las agresiones contra PDDHyP; 3) las características de las víctimas por recolectar y sistematizar; 4) el período de tiempo para analizar; y 5) las fuentes de información a partir de las cuales se recopilarían los datos.

Respecto al último punto, en principio se estableció que se recolectaría información de cuatro periódicos nacionales y 31 periódicos locales. Originalmente, se buscaba tener dos periódicos por estado[1]. Sin embargo, resultó imposible alcanzar esta meta. Los motores de búsqueda de los periódicos en sus páginas de internet son muy diversos y la manera más eficiente de realizar la búsqueda de notas fue mediante la base de periódicos de EMIS/Infolatina, la cual recientemente experimentó cambios sustanciales y redujo el número de fuentes de medios impresos disponibles en la misma. De cualquier manera, se generó la mejor selección posible de prensa, para así alcanzar una cobertura amplia y diversa para las distintas regiones geográficas del país. Adicionalmente, se estableció la inclusión de archivos de reportes, alertas y notas de monitoreo de cuatro organizaciones de la sociedad civil (OSCs) dedicadas a la defensa de PDDHyP. Se provee mayor información al respecto en la sección de Metodología más adelante, así como en el Manual que acompaña a este documento.

Tras una reunión con el Espacio de OSCs del Mecanismo para la Protección de PDDHyP, se solicitó incorporar un periódico nacional y tres periódicos locales más, así como considerar los reportes y alertas de seis organizaciones adicionales dedicadas a la defensa de PDDHyP.

II. CONSTRUCCIÓN DE DICCIONARIOS, PRUEBAS DE BÚSQUEDA Y RECOLECCIÓN DE INFORMACIÓN

Una vez que se definieron las fuentes de información, inició el proceso de recolección de información sobre eventos de agresión contra PDDHyP. Para ello, primero fue necesario definir los diccio-

[1] No se agregaron datos de las procuradurías porque no hubo acceso a dicha información y no se abrieron expedientes por la delicadeza del tema. Además, es importante señalar que no todas las agresiones que Datalab recopiló necesariamente serían denunciadas a las procuradurías. Es bien sabido que hay una enorme cifra negra en el país, en particular para este tipo de agresiones en donde las mismas autoridades son partícipes.

narios de términos que guiarían la búsqueda de notas en los periódicos, asegurando la recopilación de textos relevantes. Tras una primera ronda de búsqueda y recolección de notas, fue evidente que era necesario establecer también términos de exclusión para eliminar notas de agresiones contra defensores o periodistas que no hubieran ocurrido dentro del territorio mexicano. Además, fue importante hacer una primera eliminación de notas que hacían referencia al mismo evento. Para ello, se identificaron las notas publicadas en fechas cercanas y con titulares iguales o similares para posteriormente ser eliminadas.

Los retos se multiplicaron en el proceso de búsqueda y recolección de información a partir de los reportes y alertas de OSCs. En primer lugar, la extracción de los textos difirió sustancialmente del método implementado en el caso de los periódicos. Fue necesario identificar dentro de los sitios de internet de cada organización la ubicación particular de sus archivos relevantes. Una vez extraídas, con el propósito de eliminar duplicados, se realizó una revisión manual por nombres de PDDHyP para quienes se emitieron alertas y se confirmó también el tipo de derechos que eran relevantes para la base.

Al mismo tiempo, se desarrolló un algoritmo para recolección de textos (*web scrapping*) que nos permitió hacer una búsqueda de palabras clave y así recuperar título, texto, fuente y fecha de todas las notas identificadas como relevantes para la base de datos.

III. DESARROLLO SOFTWARE PARA LECTURA Y SISTEMATIZACIÓN AUTOMATIZADA

Siguiendo la lógica de *EventusID*[2], se desarrolló un software capaz de realizar una búsqueda automatizada de términos en las 2432 notas recopiladas y cuidadosamente seleccionadas en el paso anterior. Dentro de cada texto relevante, se examinaron cinco elementos principales: 1) nombre de la víctima, 2) tipo de víctima (periodista, activista, organización o medio de comunicación), 3) verbo que indica el tipo de agresión perpetrada, 4) lugar de ocurrencia del evento (estado y municipio) y 5) fecha de ocurrencia. Ver el apartado Conceptualización en el informe presente (página 12) o el Manual para más información.

Para la lectura y sistematización automatizada de estos campos, se crearon y refinaron nuevos diccionarios específicos. Con este propósito, se desarrolló la metodología *Word2Vec* para mejorar la lectura de notas y contextualizar los términos identificados en dicho proceso. Asimismo, para identificar los nombres, tanto de víctimas como de estados y municipios, se separaron los nombres que iniciaban con una letra mayúscula (*Ngrams analysis*).

Uno de los retos fundamentales en esta etapa incluyó la lectura de alertas por las OSCs, cuyo formato y estructura de redacción difiere sustancialmente de aquella de las notas periodísticas. Una nota periodística suele reportar un solo evento, pero los textos de la OSCs generalmente contienen información sobre diversos eventos y hacen referencia a múltiples autoridades hacia las que dirigen sus demandas. Como se señala a continuación, esto nos forzó a realizar una revisión manual de cada nota.

[2] Osorio, Javier y Alejandro Reyes. 2016. "Supervised Event Coding From Text Written in Spanish Introducing Eventus ID." *Social Science Computer Review* (doi: 10.1177/0894439315625475).

En esta etapa también fue importante la eliminación de duplicados, lo cual se hizo con base en las coincidencias de registros en cuanto a cuatro campos: 1) nombre de la víctima, 2) tipo de evento, 3) fecha y 4) lugar de ocurrencia. Desafortunadamente, los textos seleccionados no siempre fueron lo suficientemente claros o exhaustivos sobre estos cuatro elementos. Por ejemplo, algunas fuentes usan los nombres completos de las víctimas o municipios y otras usan una versión corta.

IV. VALIDACIÓN Y REVISIÓN MANUAL DE NOTAS

Como se señaló anteriormente, dada la estructura y redacción diversa de los textos recopilados, no siempre fue posible incorporar información de manera automática para todos los campos de interés de este proyecto de cada uno de los eventos recolectados. Además, algunos textos hacían referencia a múltiples eventos o víctimas y resultaba necesario identificar con cuidado las características de cada agresión referida. Por lo tanto, inevitablemente se tuvo que realizar una revisión manual de los registros con el fin de identificar plenamente la información de estos campos. Se aprovechó además para validar cada nota recopilada en la etapa anterior.

En este proceso fue claro que varias notas eran sumamente ambiguas en su reporte de víctimas (e.g. “Reporteros resultan lesionados en protesta” sin dar números concretos) o bien, que dentro de una misma nota señalaban varias agresiones que no se debían de contabilizar dos veces (e.g. “Explotan oficinas de ONG y muere activista.”). Por tanto, se concluyó que aunque se rescatarían los nombres de todas las víctimas posibles, DataLab sólo mediría el número de eventos agresión y no el número de víctimas. Dada la ambigüedad de varios de los textos, el número de víctimas es un dato que no nos sería posible validar de manera homogénea y objetiva. Esto mismo es lo que Osorio y Reyes (2016) pudieron reportar en su base de datos, e.j.. eventos mas no número de víctimas. En resumen, reportar a las víctimas de manera individual se escapó de la capacidad de nuestro software, así como de la capacidad del equipo. Sin embargo, esto no significa que, hacia el futuro, no se pueda hacer el conteo individual, ya que DataLab entregará a la Unidad de Prevención Seguimiento y Análisis los insumos necesarios para hacer tal conteo de manera detallada. Por ahora, tener el número de eventos de agresión contra PDDHyP revela información muy relevante sobre los patrones de violencia contra este grupo de personas en situación de vulnerabilidad y realizar el análisis propuesto inicialmente.

El proceso de validación manual también reveló que, en lo general, la identidad de los agresores era ambigua. Por tanto, nos fue imposible incluir esta información en la sistematización de las notas.

V. RECOPILACIÓN Y SISTEMATIZACIÓN DE VARIABLES INDEPENDIENTES

En colaboración con los equipos de la Estrategia Digital Nacional y del Mecanismo para la Protección de PDDHyP se recolectaron insumos para la construcción de la base de datos de variables explicativas en torno al fenómeno de las agresiones contra defensores y periodistas en México.

En términos generales, se recopiló y sistematizó información sobre:

- Dinámicas de conflicto territorial
- Violencia criminal
- Estado de derecho
- Respeto a los derechos humanos
- Conflictividad social
- Fortaleza de la sociedad civil y la comunidad periodística
- Características sociodemográficas

La operacionalización y medidas específicas para cada una de estas dimensiones se puede consultar en el "Manual de Documentación, Metodología y Análisis," además de lo que se describe en la sección de metodología del análisis estadístico más adelante en este documento.

VI. ANÁLISIS ESTADÍSTICO DE LAS AGRESIONES CONTRA PDDHYP

Tras unir las bases de datos correspondientes a las variables dependiente, independientes y de control, se procedió al análisis estadístico de las agresiones contra defensores y periodistas.

Metodología

A continuación se describen los rasgos generales de la metodología que en la cual se basó este proyecto. Los pormenores de cada etapa y su metodología específica se desarrollan detalladamente en el documento "Manual de Documentación, Metodología y Análisis" de la base de datos de eventos de agresión contra PDDHyP."

I. CONCEPTUALIZACIÓN

En el estudio de las agresiones contra Personas Defensoras de Derechos Humanos y Periodistas (PDDHyP) se debían definir tres conceptos básicos: 1) tipo de víctimas, 2) tipo de violencia y 3) tipo de agresiones.

Las definiciones sobre defensores de derechos humanos y periodistas se basaron en los términos de las regulaciones y declaraciones internacionales existentes en materia de derechos humanos y periodismo.

En lo que se refiere al concepto de agresiones contra PDDHyP relevante para la base de datos, se les definió como "el daño a la integridad física o psicológica, amenaza, hostigamiento o intimidación que por el ejercicio de su actividad sufran las Personas Defensoras de Derechos Humanos y Periodistas." De manera más específica, se estableció considerar solo las agresiones a aquellos derechos que protegen la integridad física o psicológica de las PDDHyP:

- Derecho a la vida;
- Derecho a la libertad;
- Derecho a la integridad personal; y,
- Derecho a la seguridad personal.

Si bien hay otros derechos importantes para el trabajo de las PDDHyP, tales como el derecho a la asociación, a la expresión, al debido proceso, a la igualdad y no discriminación, a la información, al trabajo digno, a la honra y a la dignidad personales, las agresiones a estos derechos no pueden ser atendidos dentro de esta base de datos. Por un lado, nuestra definición de agresión —basada en la Ley para la Protección de Personas Defensoras de Derechos Humanos y

Periodistas— acota el tipo de actos en los que nos enfocaremos y que claramente sólo caben en el conjunto de derechos anteriormente enunciados—i.e. vida, libertad, integridad personal y seguridad.

Por otro lado, para identificar y medir correctamente injerencias en el goce de este otro grupo de derechos sería necesario recolectar información a través de fuentes directas, encuestas y entrevistas a los actores principales, ya que se trata de un tipo de violaciones que los medios suelen no cubrir de manera consistente o bien, que las víctimas suelen no denunciar de manera abierta y consistente. A pesar de estas limitaciones, consideramos que la presencia e incidencia de agresiones al derecho a la vida, a la libertad, a la integridad y a la seguridad muy probablemente refleja otros tipos de transgresiones más complejas que van destinadas a coartar el derecho a la libre expresión y a la información, por ejemplo. Dada la posibilidad de medir y definir de manera más objetiva las agresiones de tipo físico o psicológico, se decidió enfocarse en las violaciones de derechos de este tipo.

En cuanto al tipo de violencia, se estableció que eran dos tipos los que resultaban relevantes para los propósitos de la base: 1) la violencia generada por las autoridades estatales y 2) la violencia generada por grupos criminales. Por lo tanto, la violencia de carácter personal entre particulares queda totalmente descartada en esta base de datos.[3]

[3] Ver sección de "Conceptualización" en el Manual de Documentación, Metodología y Análisis de la Base de Datos sobre Eventos de Agresión contra Personas Defensoras de Derechos Humanos y Periodistas.

II. OPERACIONALIZACIÓN

a. Agresiones y sus características

Tras definir los conceptos anteriores, se procedió a la identificación de las formas de medir cada una de sus características y dimensiones.

Partiendo de la definición anterior de "agresiones," se desarrolló un listado de las mismas según el tipo de derecho vulnerado y se especificaron descripciones para sus posibles manifestaciones, con el fin de facilitar al software la codificación de cada agresión.

Dados los tipos de violencia previamente definidos como relevantes para la construcción de la base de datos, se listaron también los posibles sujetos perpetradores de agresiones contra PDDHyP. Sin embargo, como se señaló anteriormente, dada la ambigüedad de las notas en cuanto a este rubro, resultó imposible recolectar y sistematizar este campo. De haberlo hecho, habríamos producido información incompleta y posiblemente sesgada, según la fuente utilizada.

Con el fin de entender las distintas dimensiones de las agresiones y el contexto en el que éstas ocurren, se establecieron las variables adicionales que serían relevantes para capturar otras características, tales como: nombre de la víctima, entidad y municipio de ocurrencia, fecha de los hechos, fecha de la nota, nombre de la fuente, entre otras.

b. Periodo de análisis

Entre otros objetivos, este proyecto buscaba mapear la distribución de agresiones en contra de PDDHyP tanto en términos espaciales como temporales. Por lo tanto, era necesario definir también el periodo de tiempo relevante. Ante el interés de la Unidad de Prevención Seguimiento

y Análisis de entender los factores que explican las agresiones contra PDDHyP, se acordó recopilar información tanto para el periodo anterior al estallido de violencia contra estos grupos, como durante y después del mismo, es decir, del año 2009 en adelante. Desde un punto de vista teórico, resulta necesario hacerlo así para poder entender cómo, cuándo y porqué se transformó la realidad de estos actores. Desde un punto de vista estadístico, un mayor cúmulo de información nos permite ajustar mejor nuestros modelos predictivos. Finalmente, desde un punto de vista práctico, como se muestra a continuación, las fuentes disponibles en línea —el medio más viable dado el tiempo limitado para este proyecto— no nos permiten ir más atrás en el tiempo.

c. Fuentes

Dado que parte del objetivo de esta base de datos es recopilar información histórica, se consultaron múltiples fuentes tanto para recolectar datos de agresiones como para cotejar eventos. Las fuentes básicas incluyen: notas periodísticas y archivos de notas de monitoreo de organizaciones de la sociedad civil que se dedican a atender a PDDHyP.

III. BÚSQUEDA DE INFORMACIÓN

En términos generales, el proceso de búsqueda de notas se realizó de manera semiautomatizada, con base en diccionarios de términos de búsqueda previamente establecidos. Se desarrollaron algoritmos que, a partir de las palabras claves previamente definidas en los diccionarios de términos, realizan una búsqueda en las páginas web de las fuentes definidas y guardan el texto de las notas en un archivo para su análisis posterior. Sin embargo, el proceso fue diferente para periódicos y para las organizaciones de la sociedad civil.

Para la búsqueda y recopilación de notas en periódicos se siguieron dos etapas. En una primera etapa, se utilizó la base de periódicos mexicanos de EMIS/Infolatina. En la segunda etapa, se complementó la información con algunos periódicos nacionales y regionales que no se encuentran en EMIS pero que se consideró importante incluir, según las sugerencias del Espacio de OSCs. Dada la diversidad de los motores de búsqueda, en esta etapa, la forma exacta en que se realizó la consulta para cada sitio varía y por lo tanto se desarrollaron algoritmos específicos para cada página. De los resultados de cada consulta se guardó en un archivo tipo *json*[4] el título, texto, fecha de publicación y nombre del periódico.

Para complementar la información de las notas de periódicos, también se incluyeron las alertas y acciones urgentes que algunas OSC publican en sus páginas web, en las que informan sobre agresiones y amenazas en contra de la labor desempeñada por periodistas, defensores de derechos humanos y activistas. En este caso no se realizaron búsquedas dentro de las páginas, ni se utilizaron palabras claves, sino que se tomaron todas las alertas y acciones urgentes publicadas hasta la fecha. Para cada página se diseñó un algoritmo que obtiene el título, texto, fecha de publicación y link de cada alerta, y guarda toda esta información en un archivo tipo *json*.

Este proceso resultó en la recopilación de más de 2,432 notas y reportes entre fuentes

[4] Formato de archivo para guardar datos en los que se puede etiquetar el contenido.

periodísticas y de informes o alertas de las organizaciones. Como se mencionó anteriormente, con esta primera recopilación de información, se eliminaron las notas duplicadas mediante su similitud en cuanto a titulares y fechas de publicación.

IV. SOFTWARE DE LECTURA Y SISTEMATIZACIÓN AUTOMATIZADA

Ante los miles de textos por leer y sistematizar, se creó un software capaz de hacer ambas tareas de manera automática. Este programa recorre el contenido de cada nota y sistematiza automáticamente los rasgos generales de cada nota recopilada en el paso anterior, generando una base de datos con información general de cada evento de agresión: a) el actor agredido (nombre propio de la víctima y su tipo —periodista, activista, medio de comunicación u organización—); b) el tipo de agresión (capturado mediante los verbos de acción que contiene la nota); c) la ubicación (municipio y estado) del evento; d) la fecha de ocurrencia; y, e) el identificador de la fuente de información.

La lectura y sistematización de textos mediante el software requiere necesariamente del refinamiento de diccionarios que faciliten estos procesos. Por lo tanto, se utilizó la metodología de *Word2Vec*, la cual asigna distancias a las palabras dentro de un texto, haciendo uso de redes neuronales y regresiones logísticas, con el fin de conocer el contexto general en el que suelen presentarse las palabras dentro de un texto. De esta manera, se puede identificar de manera más eficiente las palabras relevantes para la identificación de agresiones contra PDDHyP. Esta metodología permitió el desarrollo de diccionarios más finos que fueron la base para la lectura y sistematización de textos a través del software.

En esta etapa también fue importante la eliminación de duplicados, lo cual se hizo con base en las coincidencias de registros en cuanto a cuatro campos: 1) nombre de la víctima, 2) tipo de evento, 3) fecha y 4) lugar de ocurrencia. Desafortunadamente, los textos seleccionados no siempre fueron lo suficientemente claros o exhaustivos sobre estas tres características. Por ejemplo, algunas fuentes usan los nombres completos de las víctimas o municipios y otras usan una versión corta.

Como se señaló anteriormente, dada la estructura y redacción diversa de los textos recopilados, no siempre fue posible incorporar información de manera automática para todos los campos de interés en este proyecto en cada uno de los eventos recolectados. Además, algunos textos hacían referencia a múltiples eventos o víctimas y resultaba necesario identificar con cuidado las características de cada agresión referida. Por lo tanto, inevitablemente se tuvo que realizar una revisión manual de los registros con el fin de identificar plenamente la información de estos campos. Se aprovechó además para validar cada nota recopilada en la etapa anterior.

V. ANÁLISIS ESTADÍSTICO

a. Variable dependiente

Tras el proceso exhaustivo de lectura y sistematización mediante el software desarrollado para este proyecto, se generó finalmente una base de datos tipo panel que contabiliza el número de agresiones totales contra defensores y periodistas para cada municipio y cada año entre 2009

y 2016. Por tanto, la base está compuesta por 11,845 observaciones. Se estimó además el número separado de agresiones contra defensores y contra periodistas. Estas medidas constituyen nuestras variables dependientes o variables a explicar en nuestro análisis estadístico.

b. Variables independientes

Con el fin de examinar los distintos determinantes de las agresiones, se desarrollaron una serie de variables explicativas cuya lógica y mediciones se basan en la revisión de literatura realizada previamente. A continuación se describe brevemente la operacionalización de dichas variables. Para mayores detalles, favor de revisar el "Manual de conceptualización, operacionalización y metodología."

- Dinámicas de conflicto territorial: Se consideró importante tomar en cuenta el fenómeno de "megaproyectos" de desarrollo, los cuales, de acuerdo con Cisneros y Espinosa (2016), han resultado en un incremento de agresiones contra defensores de derechos ambientales. Se construyó, por tanto, una variable que mide el *número acumulado de megaproyectos de obra pública en los últimos cuatro años a nivel estatal*. Los datos se obtuvieron de la base de datos de Proyectos y Programas de Inversión Pública que registra la Secretaria de Hacienda y Crédito Público y sólo incluye los proyectos financiados con recursos del gobierno federal.[5] Para la construcción de la variable, se seleccionó aquellos proyectos que en su descripción incluían palabras relacionadas con megaproyectos (i.e. construcción presa), según la definición provista por el "Protocolo de Actuación para quienes imparten justicia en casos relacionados con proyectos de desarrollo e infraestructura" (SCJN 2014: 10).

- Violencia criminal: La literatura en represión y violaciones de derechos humanos han encontrado que los lugares con presencia de violencia criminal tienden a presentar agresiones contra periodistas (Holland y Ríos, 2015). En este proyecto se utilizó la *tasa de homicidios por cada 100,000 hombres entre 15 y 44 años cometidos con armas de fuego* debido a que las características presentadas en este grupo de edad corresponden a la variación en tiempo y espacio de los homicidios relacionados con el crimen organizado reportado por el gobierno (Calderón et al., 2015) . La tasa se calculó a partir de los microdatos de defunciones por agresión con arma de fuego del INEGI (códigos CIE X930-959) y las proyecciones de población de la Conapo.

- Estado de derecho: En la literatura sobre represión, distintos autores argumentan que un sistema de rendición de cuentas eficaz e instituciones fuertes y legítimas disminuye significativamente la probabilidad de violación de derechos humanos (Davenport y Amstrong, 2004, Bueno de Mesquita et al, 2005; Davenport, 2007). Debido a esto, se mide el sistema de garantías con dos variables diferentes. La primera variable mide la *cifra negra*, esta variable tiene como objetivo medir la confianza y fortaleza que las personas

[5] La base de datos fue suministrada por el equipo de la Coordinación de Estrategia Digital Nacional.

perciben en las instituciones. Los datos se obtuvieron de la Encuesta Nacional de Victimización y Percepción sobre Seguridad Pública ENVIPE, esta encuesta está disponible a partir del 2010 hasta el 2015 a nivel estatal. La segunda variable busca conocer la preservación de los derechos humanos y se mide mediante las *recomendaciones emitidas por la CNDH a nivel estatal*. Para ello, se utilizaron los reportes anuales de la Comisión Nacional de los Derechos Humanos (CNDH). A partir de la revisión de los reportes de 2009 hasta 2015, se construyó una base panel a nivel estatal en la cual se contabilizaron el número de recomendaciones que se hicieron a autoridades (tanto locales como federales).

c. Variables de control

- Conflictividad social: La literatura argumenta que la represión depende de la percepción de amenazas, tales como la protesta social, en particular en los lugares que presentan instituciones débiles y pocos canales para hacer válidas las demandas (Davenport, 1995; Gartner y Regan, 1996; Regan y Henderson, 2002; Von Doeep y Young, 2013; King et al., 2013). En este análisis se midió la conflictividad social a partir de dos variables. La primera mide el *número de protestas a nivel estatal*, por lo que, la variable es un conteo. Los datos que se utilizaron para esta variable se consiguieron a partir de la base de "GDELT Analysis Service." La segunda variable se refiere a los *municipios que forman parte de la ruta migrante hacia Estados Unidos*. Se esperaría que en los municipios en que transitan migrantes haya más actividad de organizaciones civiles y, como consecuencia, existan más agresiones en su contra. La variable se construyó a partir de las siete rutas de migrantes que identificó el Instituto Nacional de Migración, a raíz del mapa se realizó una intersección entre las rutas y los municipios de México. Esta variable es dicotómica y su unidad de análisis es el municipio. Por lo tanto, se codificó a los municipios que son paso para los migrantes (1) y los que no lo son (0).

- Contexto político-electoral: Se buscó medir también el nivel de democracia electoral en los municipios y estados. Las investigaciones sobre represión han encontrado que la competencia pluripartidista y una participación competitiva produce menos niveles de violación a derechos humanos (Bueno de Mesquita et al., 2005). También, se ha revelado que conforme diversas dimensiones de la democracia mejoran (i.e. prensa libre, oposición política, rendición de cuentas), entonces también la probabilidad de que el gobierno reprima disminuye significativamente (Davenport y Amstrong, 2004, Bueno de Mesquita et al, 2005; Davenport, 2007; Whitten-Woodring y James, 2012). Por estas razones, los modelos estadísticos controlan tres variables políticas-electorales. La primera variable es el *número efectivo de partidos* (NEP) que mide la pluralidad de partidos y la existencia de una oposición política. El NEP se calculó a partir de la fórmula propuesta por Laakso y Taagepera (1979). La segunda variable mide la *coordinación política* para así mostrar la capacidad que puede tener el municipio frente al estado. Siguiendo el trabajo de Trejo y Ley (2016), se trata de una variable dicotómica, donde el uno (1) representa que existe

coordinación entre municipio y estado y el cero (0) representa lo contrario. La tercera variable político-electoral mide si en un año existieron *elecciones municipales*. Siguiendo VonDoepp y Young (2013) es posible que los municipios donde hubo elecciones tengan potencial de ser reprimidos porque el gobierno o el crimen organizado quieren controlar el flujo de información en ese momento particular de incertidumbre política. La variable es dicotómica y evalúa si hubo elecciones municipales (1) o si no hubo (0). Todos los datos político-electorales se consiguieron a partir de los institutos electorales locales y del calendario electoral publicado por el Instituto Nacional Electoral (INE).

- Densidad de la sociedad civil y el gremio periodístico: Se esperaría que aquellos lugares donde existen más periodistas y organizaciones civiles haya también redes sociales más sólidas y extensas entre estos grupos, aunque también posiblemente mayor exposición de los mismos. Por consiguiente, la fortaleza de los periodistas se evalúa con la *tasa de personas que trabajan en periódicos, radiodifusoras y televisoras por cada 100 mil habitantes*. Los datos se recopilaron del Directorio Estadístico Nacional de Unidades Económicas (DENUE) a partir del 2009 hasta el 2016, a nivel municipal. La fortaleza de las organizaciones civiles mide la *tasa de organizaciones de la sociedad civil por cada 100 mil habitantes*. Esta variable se construyó a partir del Sistema de Información del registro de las OSC (SIRFOSC) de la Comisión de Fomento de las Actividades de las Organizaciones de la Sociedad Civil del 2009 al 2016, a nivel municipal y sólo se clasificó a las organizaciones que se dedican a la defensa y promoción de derechos humanos.

- Variables sociodemográficas: Se incluyeron dos variables sociodemográficas. La primera es el porcentaje de población indígena a nivel municipal. De esta manera, se busca estimar si en los territorios con mayor población indígena hay mayor probabilidad de agresiones contra periodistas y defensores de derechos humanos. Esta variable se obtuvo a partir del censo poblacional y su reporte de población de habla indígena. La segunda variable se refiere al índice absoluto de marginación a nivel municipal. Esta variable es un proxy de la condición socioeconómica del municipio. Este índice mide carencias en las dimensiones de educación, vivienda, distribución de la población e ingresos. De igual manera, esta variable da una idea sobre el contexto social de la población, si el índice tiene un valor más alto, indica que los habitantes enfrentan condiciones de vida menos favorables para su desarrollo. El cálculo del índice de marginación se obtuvo de Conapo para los años 2000, 2005 y 2010.

- Controles temporales: Dada la dinámica temporal de las agresiones y el visible incremento de éstas en el tiempo, se incluyeron variables dicotómicas para cada año de estudio.

d. Modelos estadísticos

Dado que se trata de un conteo de eventos, cuya media es menor a la varianza, se utilizó un modelo panel binomial negativo, cuya estimación toma en cuenta la naturaleza de esta variable. Seleccionamos este tipo de modelos en lugar

de modelos Poisson, que también se usan para datos de conteos, porque hay sobredispersión en los datos, es decir, la varianza de la variable dependiente es sustancialmente mayor que la media, lo que invalida los supuestos de la regresión Poisson.

La muestra sobre la que estimamos los modelos es un panel de 2007 municipios, con seis años de corte temporal. En la estimación final no se incorpora la totalidad de los 2457 municipios mexicanos porque no todos cuentan con información completa para todas las variables independientes. De igual forma, no consideramos un periodo más amplio por limitaciones en el periodo para el que está disponible la mayoría de variables. Realizamos las estimaciones teniendo en cuenta la estructura panel de los datos. Utilizamos el software Stata 14.

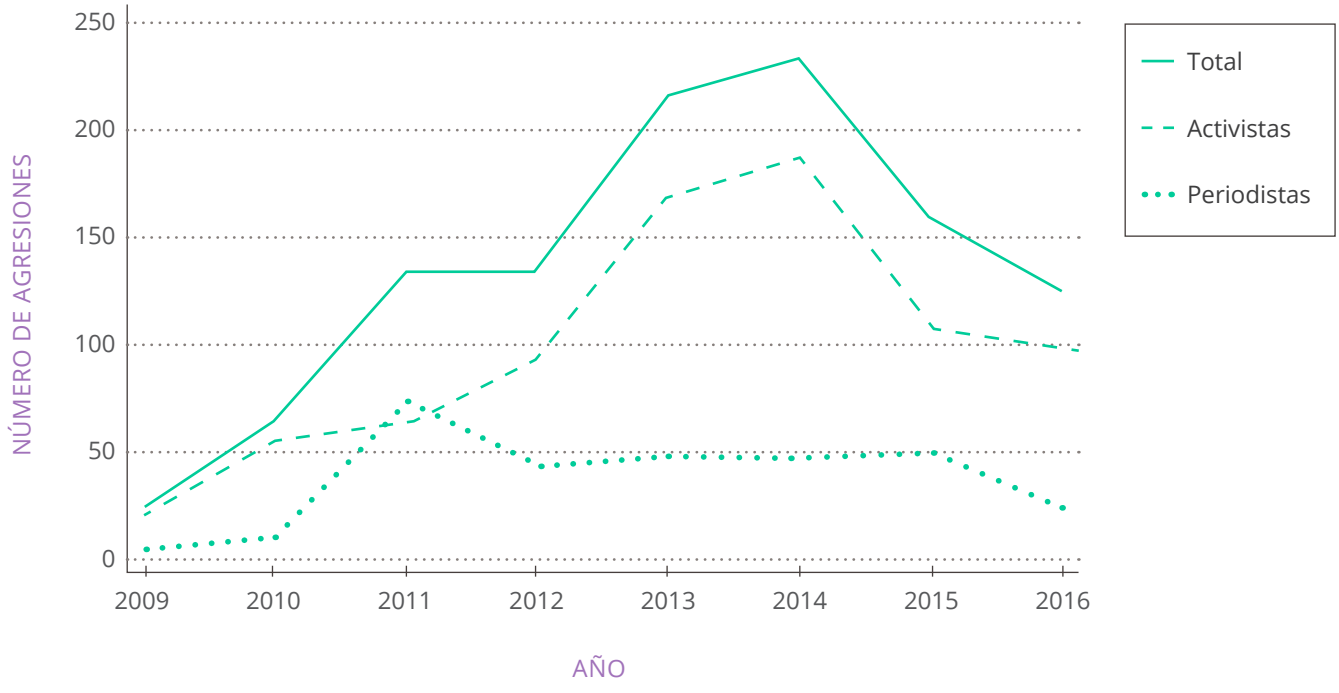
En todos los modelos incluimos a la población municipal como el *exposure* . El *exposure* se usa en las regresiones binomiales negativas para ajustar la variable dependiente como una tasa, y de esta manera tener en cuenta que al comparar el número de agresiones entre municipios se debe considerar el tamaño de la población.

Resultados

I. ESTADÍSTICA DESCRIPTIVA

La Figura 1 visualiza la evolución del número de agresiones totales contra periodistas y activistas, así como de manera separada para cada grupo. Se puede observar que la línea sólida en azul muestra el total de agresiones. El año más violento para los periodistas y activistas fue en el 2014

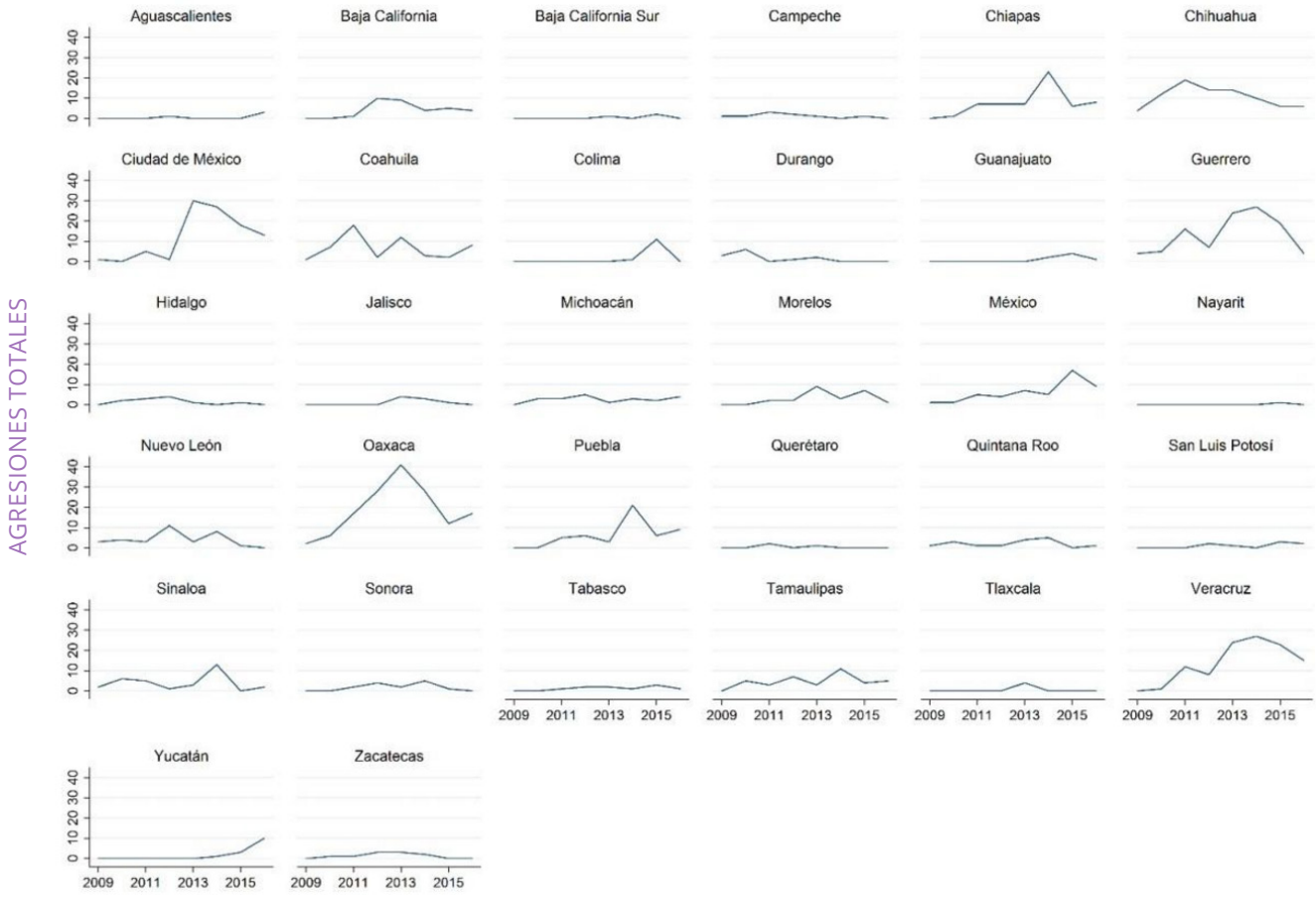
FIGURA 1. **Evolución temporal de las agresiones contra periodistas y activistas, 2009-2016**



con 233 agresiones. No obstante, cuando se desagregan las agresiones por tipo de víctima, el año más violento para periodistas no es el mismo que para los activistas. Por una parte, la línea punteada en rojo muestra las agresiones contra activistas, el año más violento para ello fue 2011, con 71 agresiones. Por otra parte, la línea punteada en verde visualiza las agresiones contra periodistas, el año más violento fue el 2014, con 187 agresiones. A pesar de que los años más violentos no coinciden para ambos grupos, en la gráfica si se visualiza que tanto el 2011 como el 2014 fueron años con un incremento en las agresiones contra periodistas y activistas.

La Figura 2 muestra la evolución de las agresiones totales contra activistas y periodistas para cada estado por separado. En los estados de Aguascalientes, Baja California Sur, Durango, Guanajuato, Hidalgo, Jalisco, Michoacán, Nayarit, Querétaro, Quintana Roo, San Luis Potosí, Sonora, Tabasco, Tlaxcala y Zacatecas presentan menos de 10 agresiones contra periodistas y activistas durante el periodo de 2009 a 2016. En el caso de Yucatán, ocurrieron 10 agresiones en el 2016, pero antes de ese año, estos eventos son prácticamente nulos. Por el contrario, los estados de Chiapas, Chihuahua, Ciudad de México, Guerrero, Estado de México, Oaxaca y Veracruz son los que

FIGURA 2. **Evolución de las agresiones contra periodistas y activistas por estado, 2009-2016**



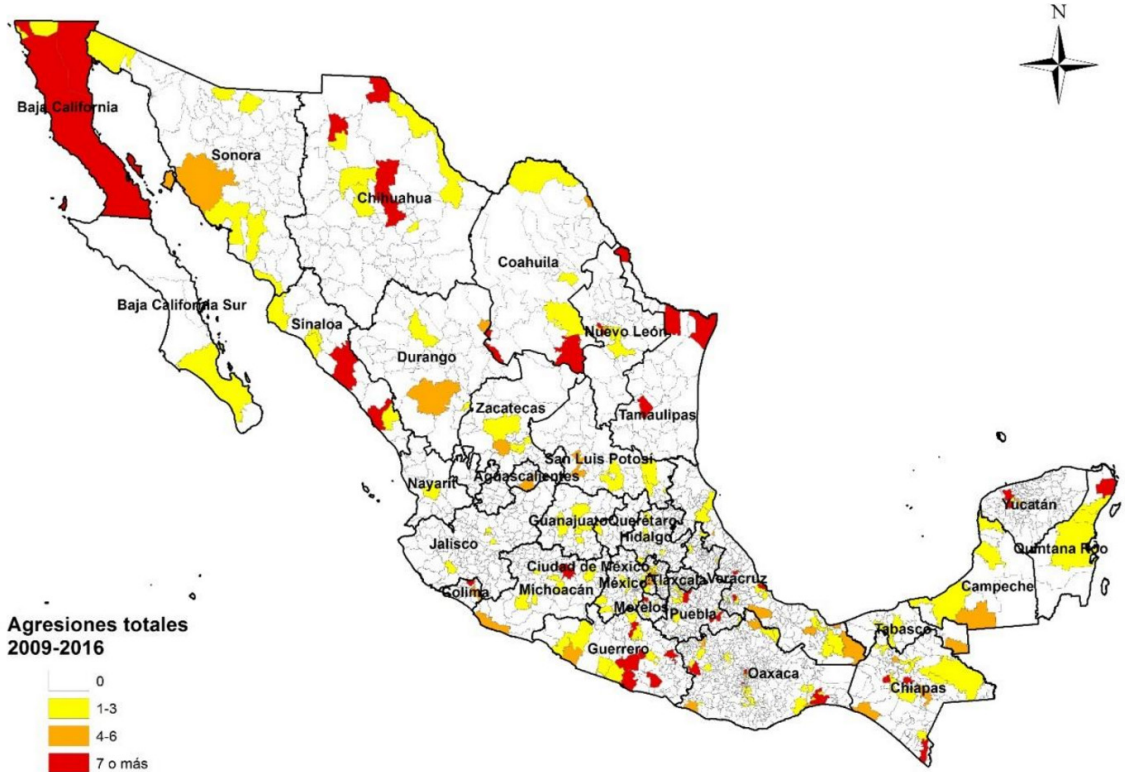
presentan más agresiones a comparación de los otros estados. En general, los estados presentan picos en los años 2011, 2013 y 2014.

Dada la variación espacial de las agresiones, el Mapa 1 ilustra esta distribución a nivel municipal. Los municipios que se encuentran en blanco son aquellos donde no hubo agresiones durante todo el periodo. Los municipios que se encuentran en rojo son los que registraron más de siete agresiones. La frontera norte y sur presentan municipios con varios eventos de agresiones. Estas regiones combinan factores de violencia criminal, paso de rutas de migrantes, entre otros posibles determinantes de agresiones que evaluamos más adelante.

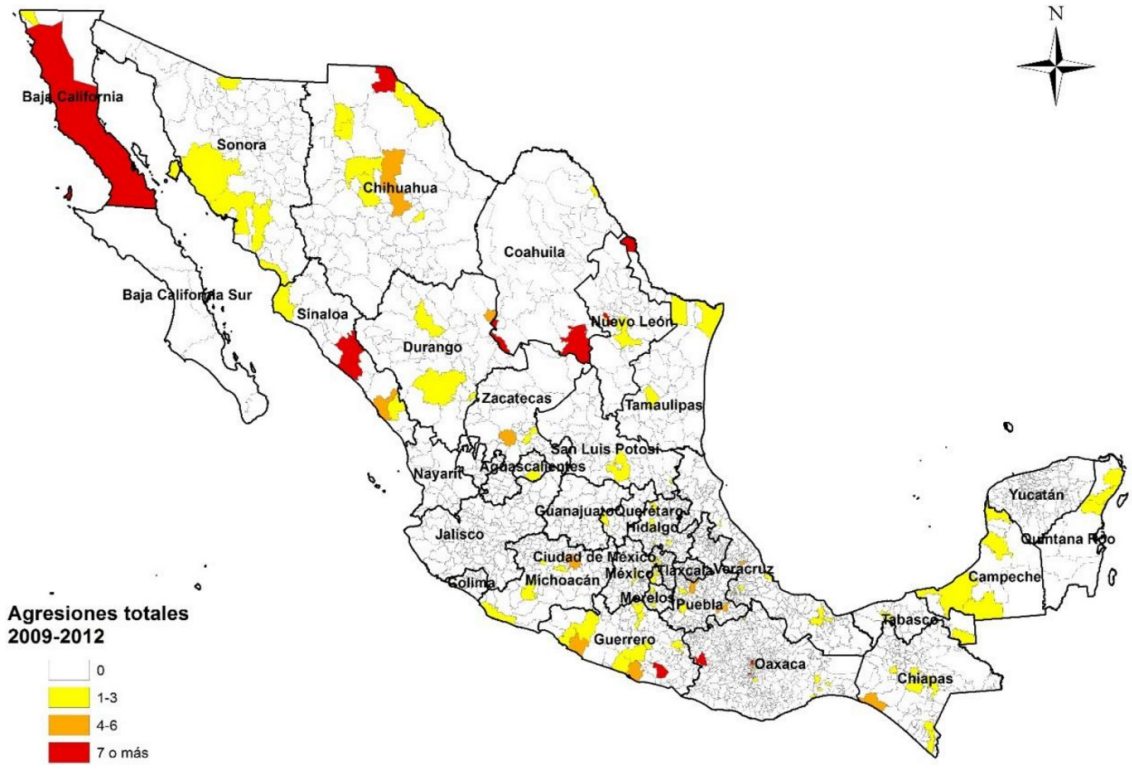
El Mapa 2 visualiza las agresiones totales sólo durante el segundo trienio del gobierno de Felipe Calderón (2009-2012). En este periodo hay municipios de la frontera norte y de la costa del pacífico con un mayor número de agresiones.

El Mapa 3 muestra las agresiones durante el primer trienio del gobierno de Enrique Peña Nieto (2012-2015). Si se compara el mapa del segundo trienio de Felipe Calderón y el mapa del primer trienio de Peña Nieto, se puede observar que las agresiones identificadas han aumentado y se han trasladado a otros municipios. El caso de Veracruz es ilustrativo, ya que durante la administración de Calderón el estado no figuraba con municipios en rojo. Para el primer trienio

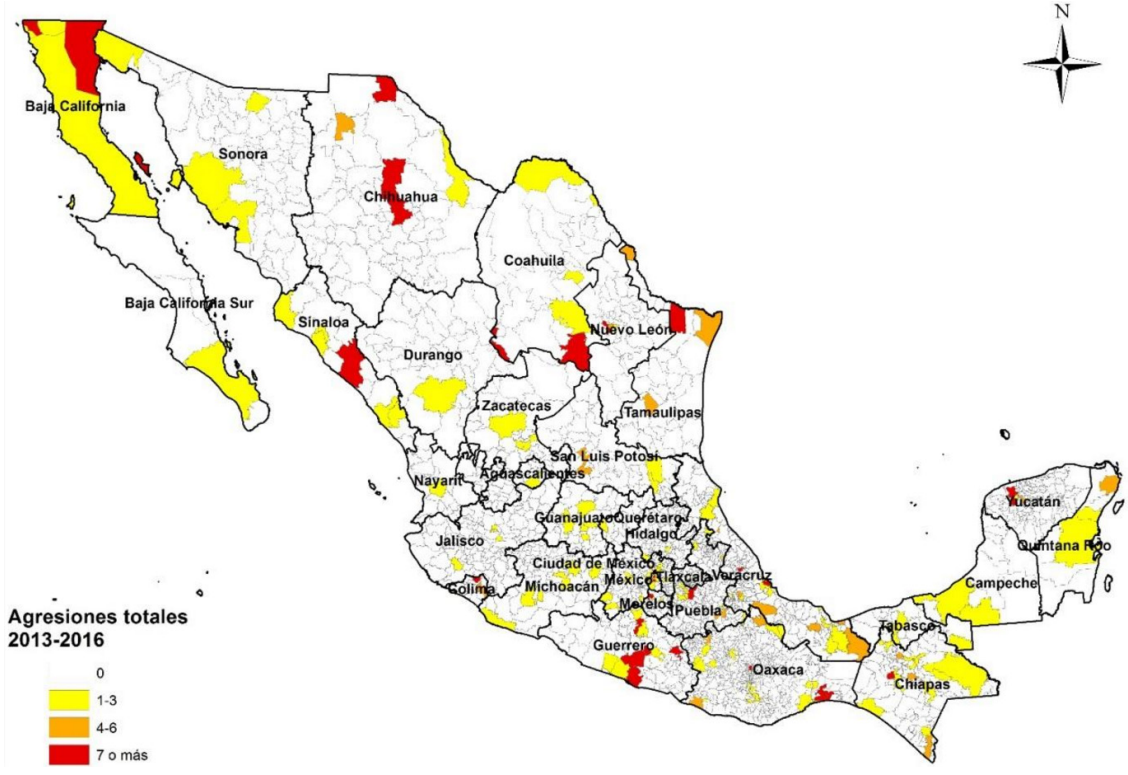
MAPA 1. **Distribución geográfica de agresiones totales a nivel municipal, 2009-2016**



MAPA 2. **Distribución geográfica de agresiones totales a nivel municipal, 2009-2012**



MAPA 3. **Distribución geográfica de agresiones totales a nivel municipal, 2013-2016**



del sexenio del presidente Peña Nieto ya es un estado con algunos municipios en amarillo, naranja y rojo. Durante este período, Javier Duarte gobernó Veracruz y sus ataques a la prensa fueron ampliamente reportados en los medios. De igual manera, Oaxaca y Chiapas son estados que tenían poco municipios coloreados durante el último trienio de Calderón, pero durante el primer trienio de Peña Nieto presentaron un incremento en las agresiones. Es importante señalar que en ambos periodos la Ciudad de México aparece coloreada. Sin embargo, la Ciudad de México al ser capital del país tiene una alta concentración de organizaciones civiles y periodistas. Asimismo, una gran cantidad de agresiones contra periodistas sucedieron durante protestas masivas, las cuales suelen ocurrir en la Ciudad de México.

II. RESULTADOS DE MODELO ANALÍTICO

La Tabla 1 muestra los resultados de los modelos estadísticos que se construyeron para identificar los factores explicativos de las agresiones contra periodistas y activistas. El Modelo 1 examina la suma total de las agresiones contra defensores y periodistas. El Modelo 2 estudia las agresiones dirigidas exclusivamente contra periodistas. El Modelo 3 se enfoca en las agresiones contra activistas.

En la Tabla se muestran tanto los coeficientes de las regresiones como el *Cociente de Tasa de Incidencia* (CTI), el cual es igual al coeficiente exponenciado y tiene una interpretación más práctica: **indica la razón en la que se incrementa la tasa de eventos al aumentar la variable independiente en una unidad**.

Es importante decir que lo que se presenta aquí es el resultado de una revisión exhaustiva de distintas especificaciones de modelos. Se exploraron los efectos de las diferentes variables explicativas de manera aislada, buscando revisar sus posibles cambios en presencia de distintos controles, así como intentando tener el periodo de estudio más amplio y el mayor número de municipios en el análisis. También se analizaron modelos que excluían a la Ciudad de México, por considerar que ésta podría tener una dinámica diferente en materia de agresiones en comparación con el resto del país---las protestas suelen concentrarse en la capital como una forma de ejercer presión entre las autoridades. Asimismo, se generaron diferentes formas de medir las tendencias en el tiempo[6]. Independientemente de estos cambios y especificaciones, los resultados del modelo principal que aquí se presenta se mantuvieron robustos.

Los resultados indican que la variable de dinámica territorial que incluimos, el número de megaproyectos de inversión pública, no contribuye a explicar el número de agresiones totales, ni de agresiones contra activistas o periodistas. Esperábamos un efecto positivo de esta variable sobre el número de ataques. Sin embargo, debemos tener en cuenta que **esta variable sólo está capturando los proyectos de inversión pública federal, dejando por fuera los grandes proyectos de inversión privada que también pueden tener un papel importante en la generación de conflictos con las comunidades**. Además, la variable está agregada a nivel estatal, por lo que tal vez no recoge adecuadamente la variabilidad de esta dimensión a nivel municipal.

[6] Se analizó cómo se comportaban las variables en presencia de *dummies* anuales y una variable continua con valores crecientes para cada año de estudio.

TABLA 1. **Agresiones contra defensores de derechos y periodistas en México, 2010-2016 (Modelos binomial negativo)**

	Modelo 1		Modelo 2		Modelo 3	
	Agresiones totales		Agresiones a periodistas		Agresiones a activistas	
	Coeficiente	CTI	Coeficiente	CTI	Coeficiente	CTI
DINÁMICAS TERRITORIALES						
Megaproyectos	0.04	1.0408	0.05	1.0512	0.0219	1.0222
	[0.0386]		[0.0435]		[0.0707]	
VIOLENCIA CRIMINAL						
Tasa de homicidios asociada al crimen	0.0014**	1.0014	0.0012***	1.0012	0.0018***	1.0018
	[0.0003]		[0.0003]		[0.0005]	
ESTADO DE DERECHO						
Cifra negra	0.0324	1.0329	0.0479	1.0491	-0.0017	0.9983
	[0.027]		[0.0309]		[0.0489]	
Recomendaciones CNDH	0.0351**	1.0357	0.0465**	1.0476	0.0371	1.0378
	[0.016]		[0.0181]		[0.0303]	
CONFLICTIVIDAD SOCIAL						
Número de protestas	0.0004	1.0004	0.0007*	1.0007	0.0001	1.0001
	[0.0004]		[0.0004]		[0.0006]	
Ruta migrante	0.4714***	1.6022	0.6230***	1.8646	0.0134	1.0135
	[0.163]		[0.1861]		[0.2754]	
CONTEXTO POLÍTICO-ELECTORAL						
Elección municipal	0.1061	1.1119	0.1408	1.1512	0.0437	1.0446
	[0.1127		[0.129]		[0.2207]	
Número efectivo de partidos	0.0486	1.0498	0.0273	1.0277	0.121	1.1286
	[0.0938		[0.1127]		[0.153]	
Coordinación partidista	-0.1695	0.8441	-0.2695*	0.7638	-0.0196	0.9806
	[0.121		[0.1387]		[0.2163]	
CONTROLES						
Tasa de OSC	0.0045	1.0045			0.0119***	1.0119
	[0.0027]				[0.0038]	
Tasa de periodistas	0.0066***	1.0066	0.0071***	1.0071		
	[0.0011]		[0.0011]			
Índice absoluto de marginación	-0.0182*	0.982	-0.0419***	0.9589	0.0031	1.0031
	[0.0096]		[0.012]		[0.0136]	
Población Indígena	0.0120***	1.012	0.0129***	1.013	0.0134***	1.0134
	[0.0034]		[0.0044]		[0.0048]	
EFECTOS TEMPORALES (año base 2010)						
2011	0.5202**	1.6824	0.0472	1.0483	1.6637***	5.2789
	[0.218]		[0.2509]		[0.4523]	

2012	0.6314***	1.8802	0.3233	1.3817	1.3823***	3.9841
	[0.2069]		[0.2283]		[0.4594]	
2013	0.8543***	2.3497	0.6627***	1.94	1.3205***	3.7452
	[0.1988]		[0.2122]		[0.4581]	
2014	0.9824***	2.671	0.8574***	2.357	1.1855**	3.2724
	[0.2192]		[0.2336]		[0.5022]	
2015	0.6308***	1.879	0.4334*	1.5426	1.2302**	3.422
	[0.2196]		[0.2309]		[0.4892]	
CONSTANTE						
	-17.6869***		-18.7130***		-16.3057***	
	[2.4861]		[2.8419]		[4.5027]	
Observaciones	11845		11845		11845	

*p<0.10, **p<0.05, *p<0.01 Errores estándar en corchetes

La variable de violencia criminal sí resultó significativa para explicar las agresiones en los tres modelos considerados, tal como lo esperábamos. Tomando como referencia el modelo de agresiones totales, el resultado indica que por cada unidad en que se incrementa la tasa de homicidios relacionada con la actividad criminal el número de agresiones se incrementa en un 0.14%. Este puede parecer un valor insignificante, pero teniendo en cuenta que el valor promedio de la tasa de homicidios en nuestra muestra es de 43.8, entonces esto implica que **entre un municipio sin homicidios y uno con la tasa promedio hay una diferencia esperada de 6.5% ([1.0014]^48) en el número de agresiones totales registradas**, manteniendo las demás variables constantes. El coeficiente en el modelo para activistas es ligeramente más alto que para el de periodistas, pero en términos prácticos se podría decir que la violencia criminal afecta a ambos actores casi por igual.

De las dos variables de estado de derecho que incluimos, sólo el número de recomendaciones de la CNDH hacia autoridades federal y locales fue significativo para explicar las agresiones totales (Modelo 1) y contra periodistas (Modelo 2). Estos resultados son consistentes con lo que esperábamos, pues indica que en aquellos estados donde hay mayores problemas con respeto a los derechos humanos, se presentan más agresiones contra los actores que pueden estar denunciando esos hechos. No obstante, se esperaba ver un efecto de las violaciones de derechos humanos sobre las agresiones a activistas (Modelo 3), pero éste no resultó estadísticamente significativo. El CTI indica que **por cada recomendación adicional que recibe un estado, las agresiones totales crecen** en 3.5% y aquellas contra periodistas aumentan en 4.7%.[7]

En cuanto a las variables de conflicto social, el número de protestas realizadas en el estado resultó ser significativa para explicar las agresiones contra periodistas (Modelo 2), aunque no para el total (Modelo 1) ni para activistas (Modelo 3). Este resultado tiene sentido ya que muchas de las agresiones contra periodistas ocurren mientras están realizando su labor cubriendo eventos de marchas

[7] Ver sección de "Análisis estadístico" en Manual de Documentación, Metodología y Análisis de la Base de Datos sobre Eventos de Agresión contra Personas Defensoras de Derechos Humanos y Periodistas

o protestas. Es importante notar que el efecto es relativamente pequeño: por cada evento de protesta se espera un aumento de 0.07% en las agresiones contra periodistas. En nuestra base registramos un promedio de 33.6 protestas cada año, por lo que la diferencia entre un estado en el que no hay protestas y uno en el que ocurre el número promedio habrá una diferencia de 2.3% ([1+0.007]^33) en las agresiones contra periodistas.

La otra variable de conflicto social, la variable dicotómica que indica si el municipio se encuentra dentro de alguna de las rutas que usan los migrantes en su camino hacia Estados Unidos, resultó ser muy significativa para las agresiones totales y para periodistas, pero no para activistas. La dirección del efecto, como lo esperábamos, es positiva: **en aquellos municipios que están en ruta de migrantes hay 60% más agresiones totales**, en promedio, que los que no están en ruta de migrantes.

Entre las variables electorales que incluimos, ninguna tiene un efecto estadísticamente significativo sobre las agresiones. La variable de coordinación entre partido estatal y municipal apenas es significativa en el modelo de agresiones contra periodistas, e indica que en los municipios en los que el partido de gobierno del estado también está presente en el gobierno municipal hay menos agresiones contra periodistas.

Las variables de control resultaron significativas en su mayoría para los tres modelos. En los municipios donde hay mayor densidad de periodistas y organizaciones de la sociedad civil ocurren más ataques contra periodistas y activistas, respectivamente.

Los resultados muestran también que a mayor rezago municipal se espera un menor número de agresiones contra periodistas. Esto puede explicarse porque los medios de comunicación suelen concentrarse en las ciudades cabecera, donde generalmente las condiciones de vida son mejores que en el resto de municipios.

Los modelos también revelan que **en los municipios con mayor proporción de población indígena ocurren más agresiones contra periodistas y activistas**. El efecto es ligeramente más alto para activistas, pero es prácticamente similar en ambos actores. Con un promedio de 19% de población indígena en cada municipio, este resultado indica que entre un municipio sin población indígena y uno con la proporción promedio hay una diferencia esperada de 25% (1.012^19) en el número de agresiones.

Finalmente, las *dummies* anuales indican que, comparado con 2010, en todos los años hubo una tendencia creciente en el número de eventos de agresión. Fue en 2013 y 2014 cuando se presentaron los aumentos más altos para periodistas, y en 2011 y 2012 para activistas.

Es fundamental señalar que, en general, los modelos para activistas tuvieron menos parámetros significativos que los de periodistas. Creemos que esto se debe en parte a que **los casos de ataques a periodistas están mejor cubiertos y predominan en la base de datos**: de los 1085 eventos de agresión que tenemos, en el 73% de los casos el actor corresponde a periodistas. Esto probablemente es así por la naturaleza del proceso de construcción de la base de datos de agresiones en el que predominó la información de medios de comunicación.

También es importante recordar que dado que el dato de la cifra negra sólo está disponible a partir de 2010, los modelos sólo analizan el periodo 2010-2016.

III. PRESENTACIÓN PLATAFORMA DE VISUALIZACIÓN

Con el fin de explotar más y mejor la base de datos, así como tener una mejor visualización de lo que ésta contiene, se construyó un tablero de visualización o *dashboard* que incluye distintos elementos y características:

- Línea de tiempo con los eventos ocurridos y nombres de las víctimas que se haya podido identificar dentro de los textos recopilados;
- Mapa de interactivo con la georreferenciación del conjunto de agresiones a lo largo del periodo de estudio; y,
- Base de datos interactiva con cada una de las variables sistematizadas en la base de datos.

Las Figuras 3 y 4 dan ilustran el contenido y tipo de visualizaciones del *dashboard*.

Se adjunta como parte de los entregables toda la documentación y archivos ejecutables para el uso del visualizador.

FIGURA 3. **Captura de pantalla de la línea del tiempo del visualizador**

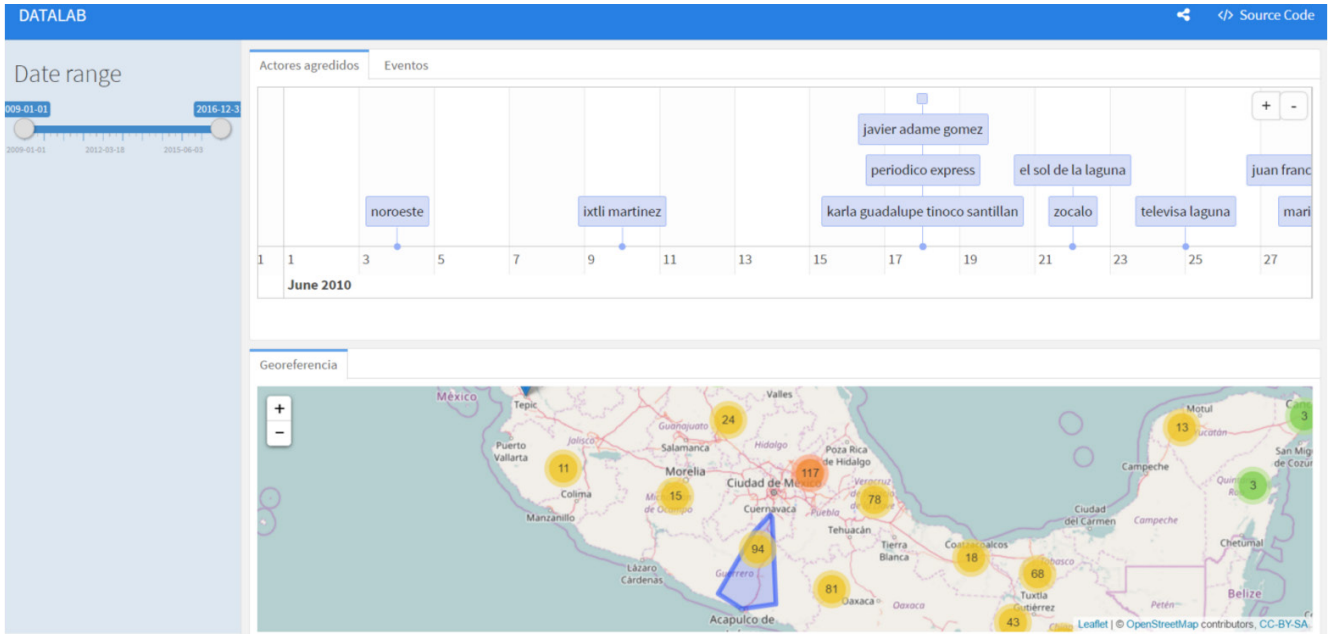
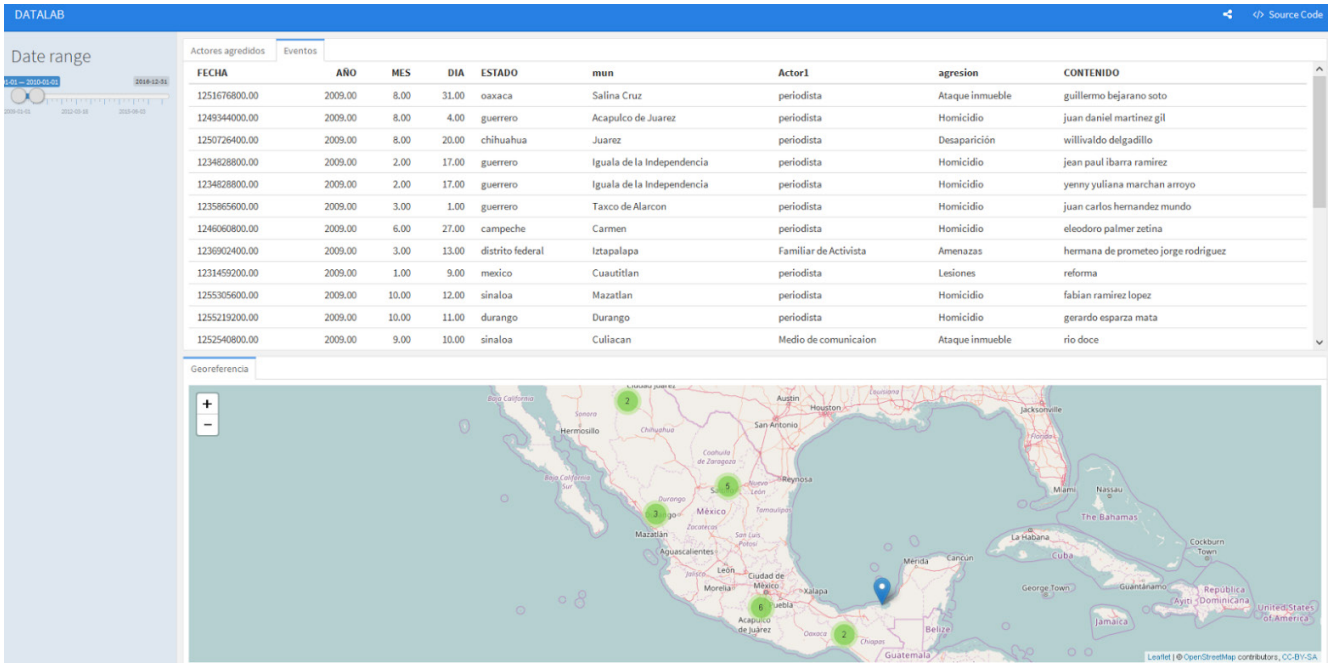


FIGURA 4. **Captura de pantalla de la base de datos interactiva del visualizador**



FECHA	AÑO	MES	DIA	ESTADO	mun	Actor1	agresion	CONTENIDO
1251676800.00	2009.00	8.00	31.00	oaxaca	Salina Cruz	periodista	Ataque inmueble	guillermo bejarano soto
1249344000.00	2009.00	8.00	4.00	guerrero	Acapulco de Juarez	periodista	Homicidio	juan daniel martinez gil
1250726400.00	2009.00	8.00	20.00	chihuahua	Juarez	periodista	Desaparición	willivaldo delgadillo
1234828800.00	2009.00	2.00	17.00	guerrero	Iguala de la Independencia	periodista	Homicidio	jean paul ibarra ramirez
1234828800.00	2009.00	2.00	17.00	guerrero	Iguala de la Independencia	periodista	Homicidio	yenny yuliana marchan arroyo
1235865600.00	2009.00	3.00	1.00	guerrero	Taxco de Alarcon	periodista	Homicidio	juan carlos hernandez mundo
1246060800.00	2009.00	6.00	27.00	campeche	Carmen	periodista	Homicidio	eleodoro palmer zetina
1236902400.00	2009.00	3.00	13.00	distrito federal	Iztapalapa	Familiar de Activista	Amenazas	hermana de prometeo jorge rodriguez
1231459200.00	2009.00	1.00	9.00	mexico	Cuautitlan	periodista	Lesiones	reforma
1255305600.00	2009.00	10.00	12.00	sinaloa	Mazatlan	periodista	Homicidio	fabian ramirez lopez
1255219200.00	2009.00	10.00	11.00	durango	Durango	periodista	Homicidio	gerardo esparza mata
1252540800.00	2009.00	9.00	10.00	sinaloa	Culiacan	Medio de comunicaion	Ataque inmueble	rio doce

Conclusión

En colaboración con la Unidad de Prevención Seguimiento y Análisis y el equipo de la Estrategia Digital Nacional, DataLab produjo una base de datos sobre el número de eventos de agresión contra los derechos de vida, libertad, integridad y seguridad de las personas defensoras de derechos humanos y periodistas. Para el periodo de análisis, 2009-2016, se registraron 1,085 eventos de agresión.

Esta base de datos permitió un novedoso acercamiento al fenómeno de la violencia que enfrentan los activistas sociales y periodistas en México, tanto para comprender su evolución en el tiempo y espacio, como para identificar algunos de los factores que influyen en su ocurrencia.

Como parte de la discusión final es importante señalar tres puntos fundamentales que deberán guiar el uso y comunicación de los hallazgos aquí compartidos.

NATURALEZA DE LAS FUENTES DE INFORMACIÓN

Se debe estar consciente que la información presentada fue recolectada a partir de notas de periódico y reportes de un subconjunto de organizaciones de la sociedad civil a las cuales tuvimos acceso en línea. Esto tiene limitaciones importantes en cuanto a las conclusiones que se pueden derivar de la base de datos y su respectivo análisis. En particular, **al utilizar fuentes periodísticas, se deben tomar en cuenta varios**

posibles sesgos o limitaciones de dichos instrumentos informativos.

Primero, puede existir cierto grado de autocensura en las notas, sobre todo en los lugares de donde hay fuerte presencia de grupos criminales. Por lo tanto, no se captura todo el fenómeno de agresiones contra periodistas.

Segundo, la recopilación de información mediante periódicos tiene tres principales desventajas: 1) sólo ciertos eventos reciben cobertura mediática; 2) los fenómenos de corta duración suelen recibir una mejor cobertura pues es menos costoso; y 3) la cobertura suele estar determinada por el espacio de páginas que haya en el periódico. En nuestro caso, el uso de periódicos locales reduce el tercer problema. De cualquier forma, *dada la naturaleza de las fuentes utilizadas, es importante tomar en cuenta que la base de datos que aquí se presenta no es ni pretende ser exhaustiva, pero sí tiene la capacidad de describir patrones de violencia contra defensores de derechos humanos y periodistas en cuanto al tipo de agresiones que se definieron como parte del estudio*.

CONTEOS DE EVENTOS VS. CONTEOS DE VÍCTIMAS

Es importante enfatizar que la base de datos presentada solamente provee información sobre eventos de agresión, pero no sobre las víctimas de dichas agresiones. Como se señala en este documento, la documentación de víctimas requiere de un trabajo mucho más detallado que resultaba imposible hacer en los cuatro meses de duración de este proyecto. Por lo tanto, enfatizamos que *esta base de datos no provee información sobre el número de víctimas afectadas en los eventos de agresión registrados en ella*.

COMPARABILIDAD CON OTRAS BASES DE DATOS DE AGRESIÓN

En esta base de datos se documentaron 1085 eventos de agresión entre 2009 y 2016. Naturalmente, la cifra de esta base difiere sustancialmente de las que otras organizaciones refieren en sus reportes. Esto se debe a que, como se describe aquí y en el Manual, en este proyecto se acotaron el tipo de agresiones a documentar. Se buscó hacer un registro que fuera lo más objetivo posible y por lo tanto, cierto tipo de agresiones como "violencia institucional" o "violencia de género" (cubiertas por organizaciones como Artículo 19) que son más delicadas y requieren un seguimiento diferente quedaron fuera del alcance del proyecto. Por consiguiente, es importante subrayar que *esta base de datos no es comparable con las bases de datos de agresiones contra personas defensoras de derechos humanos y periodistas de otras organizaciones de la sociedad civil, pero su utilidad radica en su potencial para identificar patrones de violencia dentro del tipo de agresiones delimitadas en la base*.

Siguientes pasos

RECOMENDACIONES PARA IMPLEMENTACIÓN DEL PRODUCTO FINAL

Para la mejor comprensión e implementación de la base de datos es fundamental leer cuidadosamente el "Manual de Documentación, Metodología y Análisis." Es muy importante que se conozcan y entiendan los alcances de la base. Toda la

información requerida para ello está contenida en dicho documento.

RECOMENDACIONES PARA ELABORACIÓN DE UN MEJOR PRODUCTO O UN PRODUCTO 2.0

De haber contado con un periodo más largo para desarrollar el proyecto, nos habría sido posible colaborar directamente con la Unidad de Prevención Seguimiento y Análisis en la elaboración de una base de datos detallada capaz de contabilizar el número de víctimas y no de eventos, como ocurre ahora. Y respecto a la base de datos actual, habría sido posible refinar más las herramientas para tratar que los usuarios finales programen lo menos posible, además de capacitarlos al respecto.

Un proyecto de duración más prolongada también nos habría permitido mejorar las mediciones de nuestras variables independientes. Por ejemplo, si bien se buscó implementar un cuestionario entre autoridades locales para recopilar información sobre la legislación y garantías prevalecientes en los estados para la protección de defensores de derechos humanos y periodistas, la corta duración del proyecto no permitió que los funcionarios contestaran dentro del tiempo otorgado, ni insistir personalmente para obtener sus respuestas.

RECOMENDACIONES DE POLÍTICA PÚBLICA

Los resultados estadísticos nos permiten derivar algunas recomendaciones de política pública para la protección de periodistas y defensores de derechos humanos en el país:

- La actividad del crimen organizado es claramente un factor de riesgo tanto para periodistas como para activistas. Por tanto, es vital identificar aquellos municipios donde el crimen organizado tiene la capacidad de controlar territorios y poblaciones, ya que con ello viene también el control de la información de los periodistas y de la denuncia de los activistas.

- La impunidad prevaleciente en algunos estados no sólo permite la violación de derechos de la población general, sino también de grupos específicos que tienen la capacidad de informar y denunciar dichas violaciones, como es el caso de las PDDHyP. Es crucial detectar cuidadosamente el engranaje institucional que permite el ataque del mismo estado contra dos pilares fundamentales de toda democracia.

- Los resultados del análisis estadístico sugieren que las protestas son un punto focal de riesgo para los periodistas. Los ataques en estos eventos pueden surgir por varios frentes y es importante que existan mecanismos de alerta y protección para los reporteros de eventos de manifestaciones, marchas o protestas.

RECOMENDACIONES DE CAPACITACIÓN

El correcto entendimiento de la base de datos aquí presentada, así como su análisis, requiere de un personal capacitado en las siguientes áreas:

- Introducción a diseño de investigación
- Sistematización de datos
- Visualización de datos

RECOMENDACIONES PARA PROYECTOS FUTUROS

Uno de los entregables de este proyecto incluye la compilación de todas las notas recopiladas en el proceso de recolección de información aquí descrito. Con estos insumos será posible para

la Unidad de Prevención Seguimiento y Análisis construir una base de datos de PDDHyP víctimas de agresiones que no sólo recopile sus nombres, sino también el tipo de organización a la que pertenecen, tipo de derechos que defienden, en el caso de los activistas, o la fuente que cubren, en el caso de los periodistas. Más aún, tras la identificación de estos casos específicos, se podría hacer un seguimiento más cuidadoso a través de otras fuentes de información y así también intentar identificar, en la medida de lo posible, a los agresores detrás de dichos eventos de violencia.

Dadas las limitaciones de las fuentes utilizadas, será muy importante complementar ésta y futuras bases de datos con los expedientes de periodistas y defensores que forman parte del Mecanismo de Defensa para Personas Defensoras de Derechos Humanos y Periodistas, además de información que pudieran llegar a compartir las mismas organizaciones de la sociedad civil dedicadas a representar a estos grupos.

Debido a que esta base de datos no incluye la violación de derechos laborales o violencia institucional que algunas organizaciones, como Artículo 19, si consideran, pero que no sería posible estudiar mediante notas periodísticas como en este caso, se sugiere también la elaboración de entrevistas que permitan un mejor diagnóstico de otros fenómenos de agresiones dentro del gremio periodístico.

Asimismo, es importante notar que el análisis estadístico aquí presentado incorporó las mediciones más aproximadas de los fenómenos que buscaba atender los hipótesis planteadas en principio por la Unidad de Prevención. Sin embargo, no fue posible construir para todos los casos las medidas ideales, en particular para lo que se refiere a: a) sistema de garantías y b) territorialidad. Si bien se enviaron cuestionarios a todas las autoridades locales, contrapartes de la Unidad de Prevención en los estados, con el fin de elaborar un perfil de la legislación vigente en materia de protección de PDDHyP, la tasa de respuesta de dichos funcionarios fue de menos del 50 por ciento. Hacia el futuro, se recomienda realizar una documentación exhaustiva de la evolución y perfil de la legislación en materia de protección a periodistas y activistas en los estados. Igualmente, es necesario construir mejores medidas de los conflictos de territorialidad. Si bien se tuvo acceso a las bases de datos de la Secretaria de Hacienda y Crédito Público sobre grandes proyectos de inversión pública para infraestructura, nuestra medida de “megaproyectos” omite la medición de inversiones por parte del sector privado transnacional, que otros han referido como crucial en la lucha de los pueblos indígenas y sus respectivos defensores por su tierra y recursos naturales. Por tanto, se recomienda también mejorar el acceso a bases de datos al respecto que permitan mejorar las mediciones en cuanto a grandes proyectos de inversión de infraestructura.

Evaluación

La Coordinación de Estrategia Digital Nacional diseñó la iniciativa de Datalab con la misión de impulsar políticas públicas basadas en evidencia, ciencia de datos y tecnologías digitales para la creación de capacidades en el gobierno. Establecimos una alianza con el Laboratorio Nacional de

Políticas Públicas del CIDE durante todo el proceso para asegurar la viabilidad y éxito de la iniciativa, y para contar con la experiencia técnica necesaria para que los proyectos elegidos contarán con solidez metodológica.

La EDN ha trabajado en el análisis de datos con muchas dependencias, pero nunca tan de cerca como en el proceso que se inició con Datalab, por lo que muchas veces los resultados no atendían las preocupaciones reales de éstas. Datalab trabajó directamente con tres dependencias para construir en conjunto los proyectos específicos, entendiendo los retos y problemas que fueran apareciendo en el proceso.

Se eligieron tres proyectos tras un largo proceso de identificación de actores y dependencias, así como problemáticas y capacidades que ayudaron al proceso de Datalab. Las Secretarías de Salud, de Gobernación y de la Función Pública fueron seleccionados porque contaban con los requisitos definidos por Datalab:

Se enfocaban en un problema específico o una pregunta concreta;

- Tenían un componente de análisis de datos;
- Había voluntad de alto nivel por parte de la Secretaría para implementar las propuestas;
- Contaban con un enlace fijo que pudiera dedicar tiempo al proyecto;
- Tenían bases de datos identificadas; y,
- Podían otorgar acceso directo a Datalab a estas bases de datos;

Además, los proyectos estaban vinculados claramente con los Objetivos de Desarrollo Sostenible (ODS) de las Naciones Unidas:

- El proyecto con la Secretaría de Salud buscó desarrollar un tablero de control para facilitar la toma de decisiones en políticas de salud pública. Este proyecto se vincula con el Objetivo 3: garantizar una vida sana y promover el bienestar para todos.
- El proyecto con la Secretaría de Gobernación buscó sistematizar la recopilación de información sobre agresiones contra periodistas y personas defensoras de derechos humanos, con el fin de tener más herramientas para protegerlos y trabajar hacia una sociedad justa, pacífica e inclusiva (Objetivo 16).
- El proyecto con la Secretaría de la Función Pública buscó desarrollar una herramienta tecnológica para la prevención y detección de potenciales conflictos de interés. Este proyecto también está relacionado al Objetivo 16, pues busca reducir la corrupción e institucionalizar prácticas más eficaces.

Se puede decir que todos los proyectos enfrentaron una serie de desafíos similares. El primero tiene que ver con el tiempo. Datalab se diseñó para trabajar en proyectos con una duración de tan solo cuatro meses. Esta restricción artificial tenía como propósito motivar la innovación y al mismo tiempo no permitir que proyectos que no iban a ser exitosos tardaran mucho tiempo en identificarse. Es decir, se buscó realizar pilotos y prototipos, pero no para tratar de ahondar o tratar de corregir las preocupaciones completas de cada proyecto y dependencia, de este modo se iniciaría un proceso de innovación que se buscaba fuera adoptado por la dependencia y no una entrega de un producto final.

El acceso a los datos fue una de los grandes retos a los que se enfrentaron los equipos. En

algunos casos estuvo relacionado con el acceso y permisos para trabajar con información sensible, mientras que en otras había limitantes técnicas que no permitían el uso de los datos en el estado en el que se encontraron y, en algunos casos un desconocimiento de la información que tiene cada Secretaría.

Sin embargo, podemos afirmar que todos los proyectos arrojaron resultados interesantes y de utilidad para las dependencia. El proyecto de SFP resultó en información académica y un mapeo de las sistemas de la Secretaría necesaria para desarrollar una plataforma de prevención de conflicto de interés. La herramienta desarrollada para Segob dejó una base de datos que permitirá mejorar los esfuerzos de protección de periodistas y personas defensoras de derechos humanos. Por último, el tablero de la Secretaría de Salud le permitirá dar seguimiento a una serie importante de indicadores e inició una hoja de ruta para desarrollar un tablero completo.

Todos estos proyectos fueron exitosos justamente porque había problemas concretos, datos para usar que tenían una utilidad identificada, expertos en datos y sistemas de información en pie, y voluntad para trabajar en equipo. Consideramos que estos ingredientes son necesarios para garantizar que proyectos de innovación en datos y desarrollo tecnológico prosperen en el gobierno.

Además del éxito en términos técnico de los proyectos, Datalab tenía una meta muy importante: la creación de capacidades en las dependencias. Se buscó atraer al gobierno perfiles expertos en ciencia de datos y desarrollo digital, para demostrar que hay un espacio para ellos en el gobierno. Se hizo una convocatoria abierta para atraer a los expertos en ciencia de datos y desarrollo, y en menos de una semana llegaron 110 solicitudes. De estos perfiles, 62 pasaron a una prueba para evaluar sus capacidades de trabajar con datos, reflejando lo que se estaría haciendo con las dependencias. El último filtro fue una entrevista con el LNPP y otra con la EDN. Al final, se seleccionó a 6 investigadores.

Los investigadores evaluaron el proyecto de Datalab y resaltaron varias áreas de oportunidad. En términos de la convocatoria, si bien hubo un consenso de que el proceso fue abierto y claro, todos mencionaron que la prueba era muy básica y que se debió de haber evaluado habilidades más técnicas. En términos de los proyectos, los participantes indicaron que el marco de tiempo de cuatro meses era muy corto para alcanzar las metas planteadas. También señalaron que, en futuros proyectos, se debería mejorar la planificación de los proyectos: tener más definidos los objetivos finales, tener acceso a los datos desde el inicio y estructurar de una forma más estandarizada la supervisión técnica por los profesores y científicos de datos.

Tras un cuestionario de evaluación, podemos decir que en su mayoría, los investigadores tuvieron una experiencia positiva. De los cinco investigadores que contestaron al cuestionario de evaluación, tres sienten que han aprovechado todas sus habilidades y conocimiento, uno opinó lo contrario, y otro tenía una opinión neutral. En términos generales, tres investigadores consideraron sus experiencias en Datalab como positiva o muy positiva, mientras que los otros dos lo consideraron ni positiva, ni negativa.

Datalab también buscó capacitar a los funcionarios públicos de las dependencias. En adición

a la colaboración directa en los proyectos, ofrecimos dos cursos de la Escuela de Métodos del CIDE: la Introducción a los Métodos Cuantitativos y la Introducción a los Métodos Cualitativos. Cada dependencia indicó que, en efecto, los cursos contribuyeron a la generación de capacidades dentro de la dependencia. De hecho, les hubiera gustado tomar más cursos avanzados de estadísticas; o cursos en materias de análisis de políticas públicas, georeferenciación, o manejo de bases de datos. Se considera que en futuras ocasiones el componente de capacitación puede ser más especializado y contar con un diseño específico para las necesidades de cada dependencia. Esto permitiría que los proyectos de innovación tengan un responsable dentro de la dependencia con los conocimiento para tomarlos de forma más directa.

Este proyecto se enfrentó con una serie de obstáculos no menor: acceso a datos dispersos, requerimientos de datos sensibles, complejidad técnica y poco tiempo. A pesar de esto, se puede afirmar que este proyecto ha sido muy exitoso. En tan sólo cuatro meses se construyó un *software* de procesamiento de lenguaje natural, se creó una base de datos de 1180 notas periodísticas remontándose a 2010 y, por último, se desarrolló un modelo de predicción de riesgo. Haber logrado solo uno de estos resultados habría sido un éxito, el haber logrado todos habla de un trabajo extremadamente profesional y eficaz.

Inicialmente se había planteado la posibilidad de que Datalab desarrollarse un programa para automatizar la búsqueda de notas periodísticas que el Mecanismo hace diariamente. Sin embargo, una vez que se empezó a trabajar los datos, nos dimos cuenta que era muy difícil crear este software por las propiedades de búsqueda en las páginas de los periódicos del país.

Para lograr hacer la lectura en tiempo real, se tendría que desarrollar un programa que estuviera descargando notas de múltiples periódicos que tienen estructuras digitales muy distintas y que están albergados en páginas web dinámicas. Se tendría que trabajar de forma individual en cada fuente para desarrollar un *software* a modo de cada página y que éste pudiera responder a los cambios de la misma. Aunque un software como este sería de utilidad para el Mecanismo, los beneficios del mismo no compensarían los costos de desarrollo y mantenimiento.

Sin embargo, dado que al inicio del proyecto se planteó la meta de diseñar este *software*, Datalab no fue exitoso en cumplir este objetivo tal y como fue inicialmente planteado. El equipo de Datalab debió haber hecho más esfuerzos para sensibilizar al Mecanismo de que este objetivo no se iba a alcanzar.

Otro problema importante del proyecto fue la falta de tiempo. Es improbable que con más tiempo se haya logrado el *software* inicialmente planteado, sin embargo, sí se habría logrado mejorar la portabilidad y utilidad del dashboard entregado a la dependencia. Amén de eso, el Mecanismo se mostró satisfecho con el trabajo de Datalab, la colaboración que hubo entre los distintos equipos y declaró que los entregables de este proyecto son de utilidad para su trabajo.

Este proyecto debe ser tomado como un modelo a seguir en futuras generaciones.

Referencias

Bueno de Mesquita, Bruce, George Downs, y Smith Smith Alastair. 2005. "Thinking Inside the Box: A Closer Look at Democracy and Human Rights". *International Studies Quartely* 49(3): 439–57.

Davenport, Christian. 1995. "Multi-dimensional threat perception and state repression: An inquiry into why states apply negative sanctions." *American Journal of Political Science* 39(3): 683-713.

Davenport, Christian, y David A. Armstrong. 2004. "Democracy and the Violation Human Rights: A Statiscal Analysis from 1976 to 1996". *American Journal of Political Science* 48 (3): 538–54.

Dutta, Nabamita, and Sanjukta Roy. 2009. "The Impact of Foreign Direct Investment on Press Freedom." *Kyklos* 62 (2): 239–57.

Fein Helen. 1995. "More murder in the middle—life-integrity violations and democracy in the world." *Human Rights Quarterly*. 17(1):170–91.

Franklin, James C. 2008. "Shame on you: The Impact of Human Rights Criticism on Political Repression in Latin America". *International Studies Quartely* 52(1): 187–211.Gartner y Regan, 1996;

Harff, Barbara. 2003. "No Lessons Learned from the Holocaust? Assessing Risks of Genocide and Political Mass Murder since 1955". *American Political Science Review* 97 (1): 57–73.

Holland, Bradley E., y Viridiana Ríos. 2015. "Informally Governing Information: How Criminal Rivalry Leads to Violence against the Press in Mexico". *Journal of Conflict Resolution*. OnlineFirst.

King, Gary, Jennifer Pan, y Margaret E: Roberts. 2013. "How Censorship in China Allows Government Criticism but Silences Collective Expression". *American Political Science Review* 107 (2): 1–18.

Regan, Patrick M., y Errol A. Henderson. 2002. "Democracy, Threats and Political Repression in Developing Countries: Are Democracies Internally Less Violent?" *Third World Quarterly* 23 (1): 119–36.

Trejo, Guillermo and Sandra Ley. 2016. "Federalism, Drugs, and Violence. Why Intergovernmental Partisan Conflict Stimulated Inter-cartel Violence in Mexico. *Política y Gobierno* 23(1): 9-52.

VonDoepp, Peter, y Daniel Young J. 2013. "Assault on the Fourth Estate: Explaining Media Harrasment in Africa". *The Journal of Politics* 75 (1): 36–51.

Whitten-Woodring, Jenifer, y Patrick James. 2012. "Fourth Estate or Mouthpiece? A formal Model of Media, Protest, and Government Repression". *Political Communication* 29(2): 113–36.

Anexos

A. LISTADO DE RECURSOS HUMANOS Y CONTACTOS

Nombre	Institución	Puesto
Jorge Díaz	Estrategia Digital Nacional	Project Manager
Carlos Vilalta	CIDE	Profesor Supervisor
Sandra Ley	CIDE	Profesora Coordinadora
Juan Santos	DataLab	Fellow
Allan Rosas	DataLab	Fellow
Berta Díaz	DataLab	Asistente de Investigación

B. ACTORES ADICIONALES QUE FUERON CONTACTADOS Y PROPORCIONARON AYUDA AL PROYECTO

Nombre	Institución	Puesto
Carlos Castro	Estrategia Digital Nacional	Asistencia técnica
Miguel Escalante	Estrategia Digital Nacional	Profesor Supervisor
Javier Osorio	John Jay College of Criminal Justice	Capacitación introductoria a EventusID

datalab