

## Fondo Sectorial CONACYT-INEGI

### Propuesta de tema para integrar la Convocatoria 2017 1

<b>Nombre del tema: 2</b>	Conceptualización y diseño metodológico para la extracción, procesamiento, presentación y análisis de información en seguridad pública a partir de sitios <i>web</i> y redes sociales.
<b>Importancia y utilidad del tema para el INEGI:</b>	<p>El INEGI ha plasmado en el Programa Estratégico del Sistema Nacional de Información Estadística y Geográfica 2016-2040 el reto y la importancia de “la producción de datos más oportunos con nuevas herramientas” ante “las innovaciones tecnológicas que han revolucionado el volumen, velocidad y variedad de datos disponibles”, el llamado <i>Big data</i>.<sup>1</sup></p> <p>Actualmente el Subsistema Nacional de Información de Gobierno, Seguridad Pública e Impartición de Justicia del INEGI recopila información de seguridad pública y justicia a partir de censos a instituciones gubernamentales, encuestas poblacionales y el aprovechamiento de los registros administrativos. No obstante, estas fuentes de información pueden complementarse a partir de los datos generados en sitios <i>web</i> y redes sociales que retratan diversos eventos ocurridos en el territorio nacional, con una gran cobertura geográfica, temporal y mediática.</p> <p>La cantidad masiva de datos generados en sitios <i>web</i> y redes sociales es susceptible de ser extraída, procesada y analizada con fines estadísticos a partir de las técnicas computacionales de minería de datos, aprendizaje de maquina (<i>machine learning</i>) y procesamiento de lenguajes naturales, todo ello bajo la aplicación de los controles de calidad, seguridad y confidencialidad de la información.</p> <p>Consecuentemente, los sitios <i>web</i> y las redes sociales podrían ser una fuente complementaria de suministro de información a la sociedad y al Estado con datos oportunos y pertinentes, incurriendo en los menores costos posibles a los informantes, en temas de</p>

<sup>1</sup> INEGI. Acuerdo de la Junta de Gobierno del Instituto Nacional de Estadística y Geografía por el que se aprueba el Programa Estratégico del Sistema Nacional de Información Estadística y Geográfica 2016-2040. Diario Oficial de la Federación, México.

# Fondo Sectorial CONACYT-INEGI

## Propuesta de tema para integrar la Convocatoria 2017 1

	seguridad pública a fin de coadyuvar al desarrollo nacional.
<b>Descripción general del problema a resolver:</b>	<p>En un contexto de costos crecientes para recopilar información y con exigencias de datos oportunos y de mayor desagregación geográfica, la información almacenada en sitios <i>web</i> y redes sociales se vuelve una alternativa para complementar o generar datos estadísticos en temas de seguridad pública. Actualmente su extracción, procesamiento y análisis es posible dado el avance tecnológico que han expandido las fuentes de información disponibles.</p> <p>En particular, los sitios <i>web</i> y redes sociales reportan diversos sucesos de seguridad pública a nivel nacional y local, tales como reportes de delitos del fuero común y federal, la presencia de bandas delictivas, grupos de autodefensa y policías comunitarias, “narcomantas”, ataques a grupos focalizados como periodistas, entre otros temas.</p> <p>Recientemente diversas publicaciones académicas<sup>2</sup> han comenzado a utilizar los sitios <i>web</i> para analizar temas de la criminalidad en México y en el mundo. A partir de éstas se ha generado información sobre seguridad pública y su respectivo uso en el análisis de las políticas públicas.</p> <p>Cabe resaltar que la conveniencia de emplear estos sitios para extraer, procesar y analizar información dependerá de la naturaleza de la temática de interés. Por consiguiente, un primer reto es contar con un diagnóstico sobre el tipo de información susceptible de ser extraída de la <i>web</i> y redes sociales. Posteriormente, se deberá contar con una metodología que ayude a determinar qué información en temas de seguridad pública disponible en la <i>web</i> cumple con los criterios de</p>

<sup>2</sup> Por citar algunos se encuentran: Ríos, V., Rivera, J. (2016) *What Media Coverage Can Do: Inducing Illegality in México*. Working Paper, Harvard University. Ramírez, A. (2016) *Media and Crime Perceptions: Evidence from Mexico*. Working Paper, Brown University. Chakravorty, S., Daripa, S., Saha, U., Bose, S., Goswami, S. (2015) *Data mining techniques for analyzing murder related structured and unstructured data*. American Journal of Advanced Computing, Vol. II (2), 47-54.

# Fondo Sectorial CONACYT-INEGI

## Propuesta de tema para integrar la Convocatoria 2017 1

	<p>veracidad, precisión, confiabilidad, coherencia y comparabilidad requeridos por las oficinas de estadísticas nacionales.</p> <p>Una vez definidas las características de la temática a analizar, se podrá definir el diseño conceptual, de captación, de procesamiento y de presentación de resultados, los cuales incluirán las series de actividades que permitan la generación de estadística básica cumpliendo los requisitos de rigor conceptual, confiabilidad y accesibilidad.</p>
<b>Objetivo general:</b>	Generar un marco conceptual y metodológico que identifique y valore el tipo de información en seguridad pública susceptible de ser extraída, procesada y analizada con fines estadísticos, a partir de los datos generados y almacenados en los sitios <i>web</i> y redes sociales.
<b>Objetivos específicos:</b>	<ol style="list-style-type: none"><li>1. Revisar las mejores prácticas y experiencias nacionales e internacionales sobre la extracción, procesamiento y presentación de resultados a partir de sitios <i>web</i> y redes sociales en el tema de seguridad pública.</li><li>2. Identificar y clasificar el abanico de temas de seguridad pública susceptibles de ser analizados en los sitios <i>web</i> y redes sociales, definiendo su calidad estadística, referencia temporal, periodicidad de la captación y actualización, desglose geográfico, unidad de análisis, población objeto de estudio, clasificadores, metadatos, fuentes de consulta, confiabilidad de los estimadores y niveles de precisión.</li><li>3. Generar una metodología que permita evaluar la calidad de la información en seguridad pública susceptible de ser extraída, procesada, presentada y analizada que cumpla con la granularidad, frecuencia y cobertura requerida para transformar en información estadística el gran volumen, velocidad y variedad de datos (<i>Big data</i>) disponibles en sitios <i>web</i> y redes</li></ol>

**Fondo Sectorial CONACYT-INEGI**  
**Propuesta de tema para integrar la Convocatoria 2017 1**

	<p>sociales.</p> <p>4. Generar una metodología que permita definir los instrumentos o métodos computacionales a emplear para la extracción, procesamiento y presentación de la información en seguridad pública disponible en sitios <i>web</i> y redes sociales, considerando las condiciones tecnológicas y la calidad definida previamente en la metodología desarrollada para este proyecto.</p> <p>5. Elaborar el diseño conceptual, de captación, de procesamiento y de presentación de resultados de algún tipo de información en materia de seguridad pública identificada previamente en la metodología (estudio de caso).</p>
<b>Tiempo estimado de realización:</b>	12 meses
<b>Producto(s) esperado(s):</b>	<p>1. Revisión de las mejores prácticas y experiencias nacionales e internacionales sobre la extracción, procesamiento y presentación de resultados a partir de sitios <i>web</i> y redes sociales en el tema de seguridad pública.</p> <p>2. Reporte de diagnóstico que identifique y clasifique el abanico de temas de seguridad pública susceptibles de ser analizados en los sitios <i>web</i> y redes sociales, definiendo su calidad estadística, referencia temporal, periodicidad de la captación y actualización, desglose geográfico, representatividad, unidad de análisis, población objeto de estudio, clasificadores, metadatos, fuentes de consulta, confiabilidad de los estimadores y niveles de precisión.</p> <p>3. Metodología que permita evaluar la calidad de la información en temas de seguridad pública susceptible de ser extraída, procesada, presentada y analizada que cumpla con la granularidad, frecuencia y cobertura requerida para transformar en información estadística el gran volumen, velocidad y variedad de datos (<i>Big data</i>) disponibles en sitios <i>web</i> y redes sociales.</p> <p>4. Metodología que permita definir los instrumentos o métodos computacionales a emplear para la extracción, procesamiento y presentación de la información en seguridad pública</p>

## **Fondo Sectorial CONACYT-INEGI**

### **Propuesta de tema para integrar la Convocatoria 2017 1**

disponible en sitios *web* y redes sociales, considerando las condiciones tecnológicas y la calidad requerida en la generación de datos estadísticos.

5. Diseño conceptual, de captación, de procesamiento y de presentación de resultados de algún tipo de información en materia de seguridad pública identificada previamente en el reporte de diagnóstico (estudio de caso). Deberá seguir los criterios especificados en la Norma Técnica para la Generación de Estadística Básica del INEGI [http://w-webinterpro18/directorios/Documentos/NormaTécnica\\_GeneraciónEstadísticaBásica.pdf](http://w-webinterpro18/directorios/Documentos/NormaTécnica_GeneraciónEstadísticaBásica.pdf) y en la metodología diseñada para este proyecto.
6. Un taller presencial dirigido al equipo del INEGI donde se exponga el desarrollo y resultados de la investigación, así como el uso de los materiales generados.
7. Reporte global de los resultados.
8. Un artículo de investigación con calidad suficiente para ser dictaminado favorablemente para publicar en *Realidad, Datos y Espacio Revista Internacional de Estadística y Geografía* <http://rde.inegi.org.mx/> , en el que se presenten de manera clara y sintética los resultados de la investigación.

#### **Observaciones:**

Se espera que la extracción de datos de los sitios *web* y redes sociales cumpla con los términos y condiciones de los sitios seleccionados y/o contar con la respectiva autorización.

En la medida de lo posible, incorporar el enfoque de género a los resultados del proyecto.