

# Big data y acceso a la información en México

**Daniel Berliner** (London School of Economics and Political Science)

**Brian Palmer-Rubin** (Marquette University)

**Jésica E. Tapia Reyes** (Investigadora Independiente)

**Benjamin E. Bagozzi** (University of Delaware)

**Aaron Erlich** (McGill University)



# Contenido

<b>Resumen ejecutivo</b>	<b>2</b>
<b>Introducción</b>	<b>3</b>
<b>Entender las solicitudes de información</b>	<b>6</b>
<b>Entender las respuestas</b>	<b>16</b>
<b>Sesgo de Expertise</b>	<b>25</b>
<b>Conclusiones y recomendaciones</b>	<b>28</b>
<b>Referencias</b>	<b>31</b>
<b>Equipo de investigación</b>	<b>32</b>



# Resumen ejecutivo

- El uso de la información ha crecido con el tiempo en respuesta a una gran variedad de necesidades.
- Mayor profesionalización de ciudadanos y miembros de la sociedad civil que la solicitan; desarrollo de un ecosistema de expertos en transparencia y rendición de cuentas.
- Aumento de respuestas engañosas por parte de los sujetos obligados, sobre todo para los solicitantes no expertos.
- *Sugerencia de política pública*: mejorar el sistema para rastrear los textos de las respuestas en formatos más accesibles.
- *Sugerencia de política pública*: mejorar la capacitación en las Unidades de Transparencia para estandarizar las prácticas para contestar a solicitudes de información pública.
- *Sugerencia de política pública*: fortalecer los lazos del INAI con las organizaciones de la sociedad civil y enlistar a estas últimas como intermediarias para apoyar el uso ciudadano del sistema.
- Para más información, ver la página web del proyecto: [bigdataytransparenciamx.lse.ac.uk/](http://bigdataytransparenciamx.lse.ac.uk/)



# Introducción

Los sistemas de acceso a la información tienen que cumplir al menos dos funciones. La primera es garantizar el *acceso masivo*, es decir, hacer cumplir el derecho de acceso a la información pública de todos los ciudadanos. Bajo esta función, cualquier ciudadano puede acceder a información básica sobre sus interacciones rutinarias con el gobierno. Dicha información puede facilitar ciertos trámites (como sacar un pasaporte o acceder al seguro de desempleo), o satisfacer las necesidades de información de los ciudadanos que quieren conocer el desempeño de los servicios públicos que le corresponden, por ejemplo, de las escuelas o clínicas públicas cercanas.

La segunda función de los sistemas de acceso a la información pública tiene que ver con el *acceso de expertos*. Esta función brinda información a actores sociales profesionales que pretenden utilizarla con fines de acción colectiva e incidencia. Aquí, las organizaciones de la sociedad civil (OSC), los partidos políticos, los periodistas y otros intermediarios son actores clave. Con base en su experiencia en las distintas áreas de política pública, el funcionamiento interno de las burocracias y la normatividad que rige la transparencia, esos actores facilitan un “ciclo de acción”<sup>1</sup> en donde la información pública es sólo un ingrediente para incidir en la realidad social de ese grupo.

Desde sus inicios, el sistema mexicano fue diseñado para dar prioridad a la primera función —el acceso masivo—pero en la práctica ha evolucionado para cumplir más efectivamente con la segunda función—el de acceso de expertos. Hoy en día, el Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales (INAI) sigue promoviendo la visión del acceso masivo con medidas como el Plan DAI y las Caravanas por la Transparencia, campañas que buscan ampliar el uso del sistema de solicitudes de información hacia los grupos más vulnerables. Sin embargo, de acuerdo con nuestros datos, la propia evolución del sistema ha provocado que el acceso a la información mediante solicitudes sea un proceso cada vez más complejo para el ciudadano común, y en cambio, las probabilidades de éxito de los usuarios más calificados han aumentado conforme sus solicitudes se han vuelto más complejas y específicas.

---

<sup>1</sup> Ver Fung, Graham y Weil (2007).





Para entender en la práctica el funcionamiento de este sistema, nuestro grupo de investigación se dio a la tarea de sistematizar y analizar el conjunto de solicitudes de acceso a la información dirigidas a sujetos obligados de orden federal e ingresadas al sistema del INAI entre 2003 y 2019. Analizamos las solicitudes y sus respuestas con técnicas novedosas de Big Data, modelos de inteligencia artificial (*machine learning*) y análisis de texto (*topic modeling*).

Nuestra investigación utiliza datos disponibles al público de la base de datos del INAI, INFOMEX, y de la Plataforma Nacional de Transparencia que la reemplazó, para los años 2003-2019. Cabe señalar que sólo usamos datos de solicitudes de acceso a la información, y no de solicitudes de datos personales (textos que el INAI no hace públicos, con justa razón). También limitamos nuestro análisis a las solicitudes de información presentadas ante sujetos obligados del gobierno federal, es decir, se excluyen las solicitudes presentadas ante entidades no federales, como los gobiernos estatales y locales, las entidades legislativas y judiciales, los partidos políticos, los sindicatos y otras. Aunque esas otras entidades se han ido incorporando gradualmente a las bases de datos del INAI, nuestro objetivo es analizar un conjunto coherente de entidades durante el periodo completo de estudio para sacar conclusiones más confiables con respecto a cambios en las solicitudes ciudadanas y en las respuestas de las dependencias de gobierno. Por último, sólo analizamos solicitudes presentadas en México, y no en otras partes del mundo, pues nos interesan las relaciones entre Estado y sociedad en México.

Muchos de nuestros métodos de investigación se enfocan en los textos de las propias solicitudes y en los de sus respuestas. Las bases de datos del INAI que contienen esos detalles importantes están disponibles al público; este estudio utiliza exclusivamente la información disponible al público. Para los textos de las solicitudes, usamos el campo principal y el campo de "otros datos" que llena el solicitante, además de cualquier archivo adjunto que el solicitante haya cargado (mismos que tenemos digitalizados). Para las respuestas, descargamos y digitalizamos de las bases de datos del INAI cerca de 2.4 millones de archivos individuales de solicitudes que tuvieran archivos públicos de respuesta asociados. Tanto para solicitudes como para respuestas, nuestros métodos de investigación usaron varios tipos de *machine learning*, como el modelado de temas y la predicción supervisada, con base en variables codificados por expertos. Entrenamos a nuestros algoritmos de *machine learning* con una muestra de 4,925 solicitudes, que fueron codificadas por un grupo de seis investigadores expertos en la política pública mexicana en 2019. Las solicitudes y sus respectivas respuestas se codificaron según 22 características distintas. Posteriormente, los resultados se usaron como datos de entrenamiento para sacar predicciones sobre toda la base de datos de las solicitudes de información y sus respuestas, lo que permitió hacer análisis exhaustivos y sumamente detallados.



En las siguientes secciones, explicamos los métodos con más detalle cuando sea relevante para las medidas que presentamos y discutimos.

Como resultado de nuestro análisis, encontramos varias razones para ser optimistas respecto al funcionamiento del sistema, aunque también identificamos algunas oportunidades para mejorar. El nivel de inserción del sistema de acceso a la información en México ha sido sumamente exitoso en las relaciones Estado-sociedad civil, sin embargo, su desempeño se ha visto limitado porque, por un lado, a pesar de los intentos del INAI por establecer un sistema de acceso a la información fácil para todos los ciudadanos, aún no es del todo funcional para atender todas las necesidades de información pública y, por el otro, porque aún no logra consolidarse como un canal de comunicación efectivo entre los sujetos obligados y la sociedad civil.

También subrayamos el potencial de las herramientas de *machine learning* para ayudar a entender mejor cómo se ha desarrollado el acceso a la información en México y la capacidad de respuesta de los funcionarios públicos, tanto a través del tiempo como entre los sujetos obligados. Estamos totalmente a favor de promover un mayor uso de herramientas similares por parte del mismo INAI y de las organizaciones de la sociedad civil en México. Nuestra investigación está informada por evaluaciones y análisis importantes que han llevado a cabo organismos como el INAI, Métrica de Gobierno Abierto, Fundar, Artículo 19, SocialTIC, y otros durante las últimas dos décadas.<sup>2</sup>

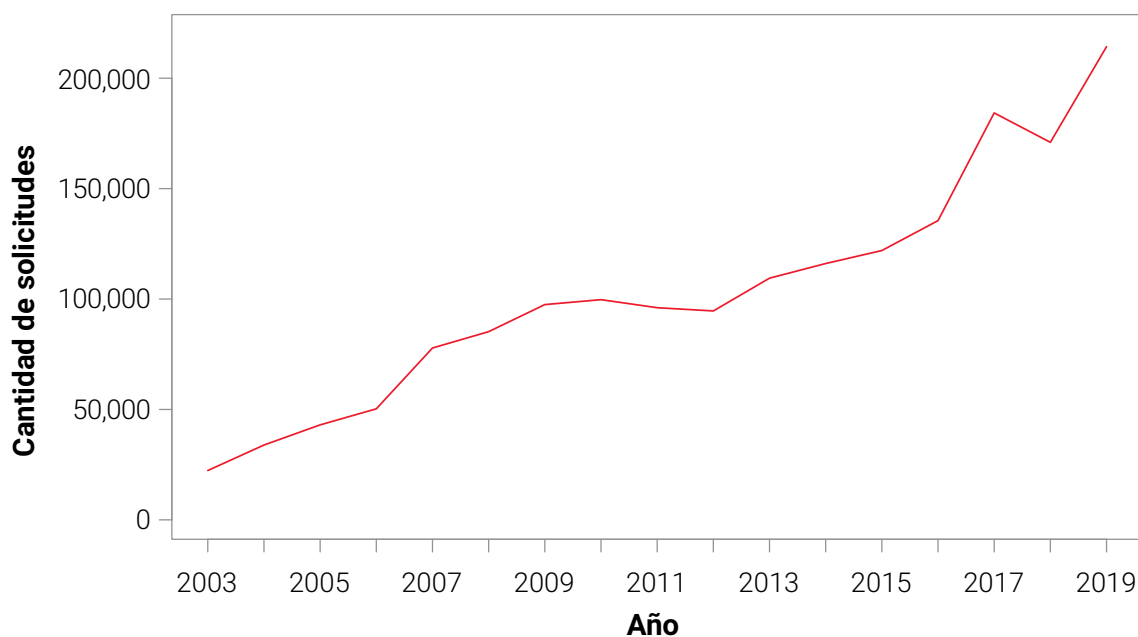
El resto del reporte se organiza en dos secciones principales. Primero, analizamos patrones en las solicitudes de información pública a través del tiempo: quién hace la solicitud y qué información busca. Segundo, observamos las respuestas a dichas solicitudes de información, presentamos nuestras nuevas estrategias para medirlas y examinamos tendencias a través del tiempo. Concluimos con un conjunto de recomendaciones de políticas públicas para seguir consolidando la figura del INAI como institución fundamental de rendición de cuentas de abajo hacia arriba en México.

---

<sup>2</sup> Ver, por ejemplo, Cejudo et al. (2014); Fox et al. (2011); Lagunes y Pocasangre (2019); y Publicaciones INAI.

# Entender las solicitudes de información

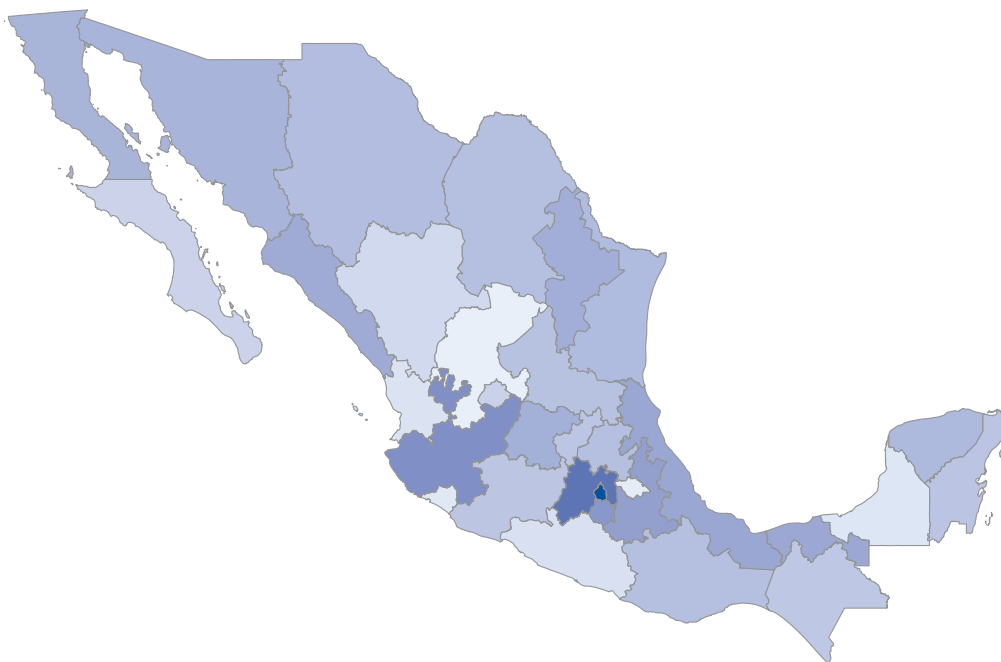
De inicio, revisamos las solicitudes de información gubernamental. Si el sistema de acceso a la información en México funciona de manera efectiva, deberíamos ver que mucha gente lo usa para fines muy diversos. En términos simples, el sistema de acceso a la información en México ha sido un éxito. Como demuestra la Gráfica 1, la cantidad total de solicitudes de información a las dependencias de gobierno federales ha aumentado año con año: en 2019 alcanzó las 200,000 solicitudes. Esto quiere decir que, desde 2004 – el primer año completo en que el sistema estuvo en operación – las solicitudes se quintuplicaron.



**Gráfica 1.** Cantidad total de solicitudes a través del tiempo (como con todos los datos de este reporte, sólo se cuentan las solicitudes a sujetos obligados federales presentadas dentro de México).

Podemos empezar a entender más sobre el uso del sistema de acceso a la información si desagregamos las solicitudes por ubicación geográfica. ¿De dónde viene la mayoría de las solicitudes de información? La Gráfica 2 desglosa la cantidad total de solicitudes por estado durante los años del programa (2003-2019); entre más oscuro sea el tono, mayor cantidad de solicitudes por estado. No es de sorprender que la Ciudad de México tenga la mayor cantidad de solicitudes a sujetos obligados federales, con un total de 809,701 (51.0% de todas las solicitudes). En segundo y tercer lugar, los estados que más información solicitan son el Estado de México (11.8%) y Jalisco (4.3%). Cabe mencionar que son los tres estados más poblados de México.

La Gráfica 3 hace un ajuste por población y muestra los estados con la mayor cantidad de solicitudes per cápita. En ese caso, la Ciudad de México sigue en la cima, con 9,148 solicitudes anuales por cada 100,000 habitantes. Resulta lógico que la capital del país esté sobrerrepresentada, dado que concentra una gran cantidad de organismos públicos, organizaciones de la sociedad civil, corporaciones y agencias periodísticas. Si observamos más allá de la Ciudad de México, los siguientes estados con mayor cantidad de solicitudes per cápita son Baja California Sur, Morelos y Tabasco.



**Gráfica 2.** Cantidad total de solicitudes entre 2003 y 2019. Un tono más oscuro indica más solicitudes.



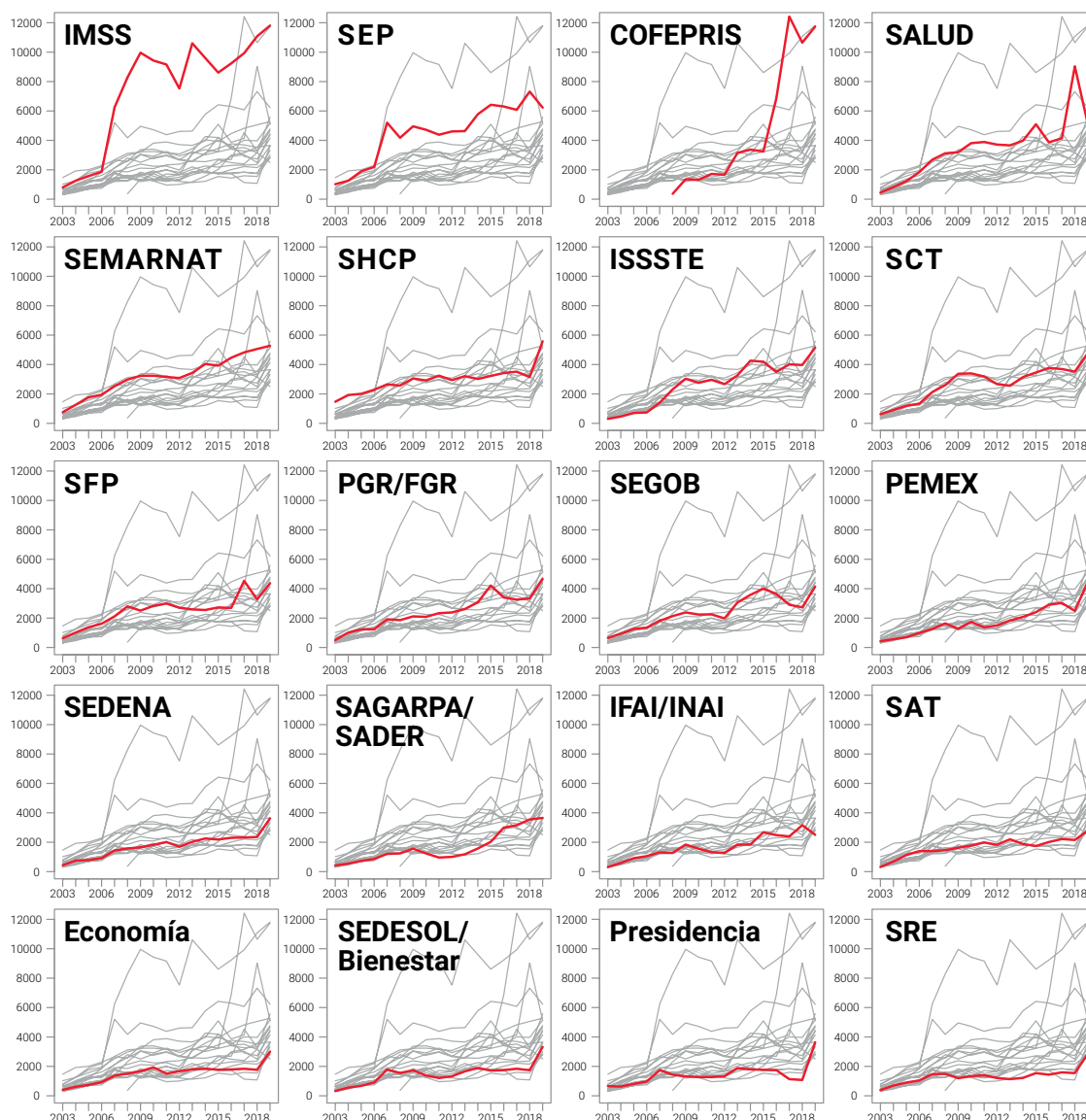


**Gráfica 3.** Cantidad de solicitudes per cápita entre 2003 y 2019. Un tono más oscuro indica más solicitudes per cápita.

Ahora pasamos a la cantidad total de solicitudes presentadas ante distintas dependencias del gobierno federal. La Gráfica 4 despliega la cantidad anual de solicitudes de información recibidas en 20 de las dependencias más interpeladas en la historia del IFAI/INAI.<sup>3</sup> El Instituto Mexicano del Seguro Social (IMSS) ha recibido más solicitudes que cualquier otra dependencia, con un total de 132,300 durante el periodo de estudio. El IMSS ha sido la dependencia con más solicitudes durante la mayoría de los años del sistema; sólo la COFEPRIS (Comisión Federal para la Protección contra Riesgos Sanitarios) recibió más solicitudes de información en 2017. Las dependencias que han recibido la mayor cantidad de solicitudes de manera consistente son la Secretaría de Educación Pública (SEP), la Secretaría de Salud, la Secretaría de Medio Ambiente y Recursos Naturales (SEMARNAT) y la Secretaría de Hacienda y Crédito Público (SHCP).

---

<sup>3</sup> Todas las dependencias aquí presentadas estuvieron entre las 20 con más solicitudes, excepto por Presidencia y la SRE, que estuvieron en el lugar 21 y 22 respectivamente. Para incluirlas, omitimos a la CONAGUA y a la CFE. Para más información sobre la cantidad de solicitudes a los sujetos obligados, véase nuestro sitio web.



**Gráfica 4.** Solicitudes a través del tiempo para cada una de las 20 dependencias seleccionadas. Las gráficas individuales van en orden descendente por cantidad total de solicitudes recibidas, 2003-2019.

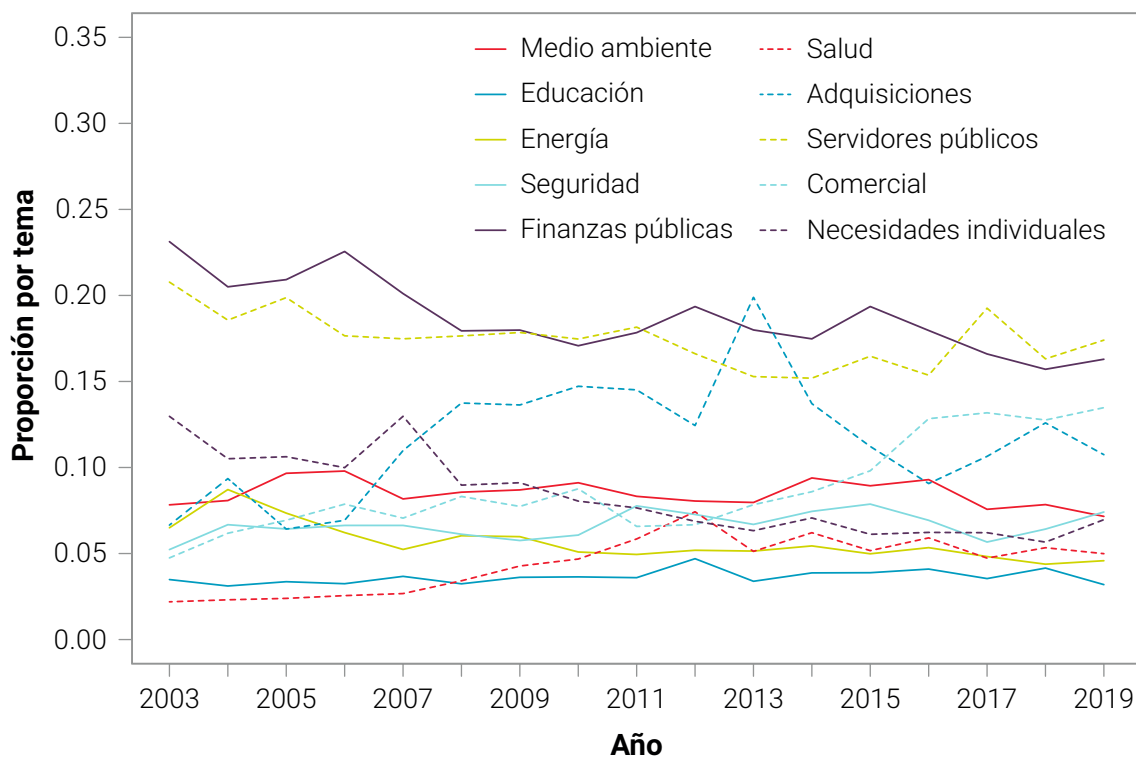
Posteriormente, tratamos de entender el tipo de información que buscan los ciudadanos. Las Gráficas 5 y 6 desglosan las solicitudes de información en dos categorías distintas. Las solicitudes de información se codificaron en dichas categorías mediante *machine learning*; a cada una se les asignó una categoría con base en el texto de la solicitud. En ambos casos, vemos que las solicitudes de información cubren de manera consistente una amplia gama de tipos de información.

Para categorizar las solicitudes por temas, usamos un método de análisis de texto llamado *unsupervised topic modeling* (modelado de temas no supervisado). Este método agrupa textos por temas con base en frases o conjuntos de palabras que tienden a aparecer juntas. Por ejemplo, palabras como “proyecto”, “agua”, “terrenos”, “forestal”, “minera” o “ambiental” indican que se trata de un tema ambiental; mientras que palabras como “policía”, “detenidos”, “militar”, “víctimas”, “delincuentes” o “armas” indican un tema de seguridad.<sup>4</sup> Aunque categorizamos las solicitudes en 20 temas distintos, para fines de este estudio combinamos varios de ellos que son muy similares para quedar con solo 10 temas.

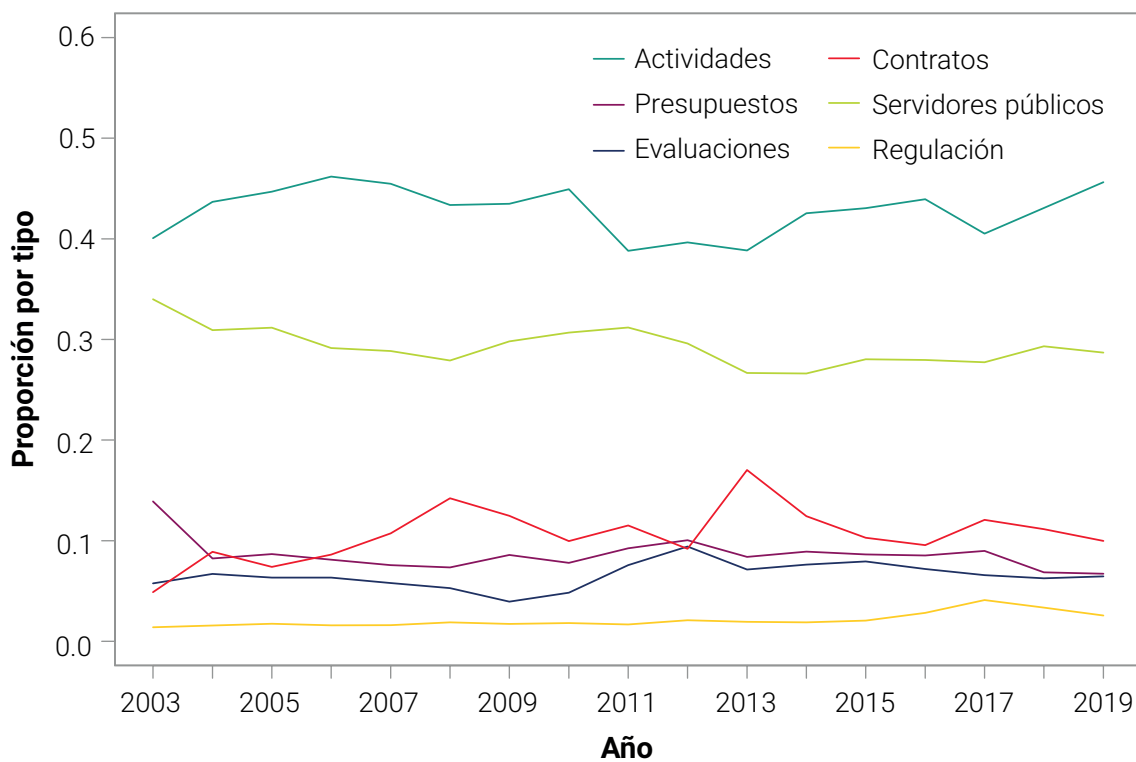
La Gráfica 5 desglosa las solicitudes de información por temas de política pública. Esto refleja un alto nivel de dispersión entre varios temas: ocho de los cuales representan de manera consistente más del 5% del total de solicitudes. Los temas más frecuentes durante todos los años del sistema son “Financiamiento público” (que tiene que ver en especial con presupuestos, gastos, impuestos y programas distributivos) y “Servidores Públicos” (que tiene que ver con contrataciones, pensiones, capacitaciones y recursos humanos). Sin embargo, ambas categorías muestran una ligera caída con el tiempo. Las categorías que muestran el mayor aumento a través del tiempo son “Comercial” (relacionada sobre todo con aprobaciones regulatorias, inspecciones, propiedad intelectual y licencias de negocios) y “Adquisiciones” (relacionada con contratos y licitaciones públicas). Al mismo tiempo, el tema que presentó la mayor *caída* a través del tiempo fue “Necesidades individuales” (relacionada con varios asuntos específicos para el solicitante, como aplicar a programas de gobierno o hacer trámites). Ese tema representó alrededor del 13% de las solicitudes en 2003, pero ha ido cayendo constantemente y, en 2019, representó sólo alrededor del 7%. Una de las posibles explicaciones de ese declive es que, conforme los sitios gubernamentales se han ido desarrollando mejor, los ciudadanos han encontrado fuentes alternativas para buscar ese tipo de información y, por tanto, han acudido menos al sistema de solicitud de información.

---

4 Para más detalles sobre los métodos, véase nuestro artículo de 2018, “What information do citizens want? Evidence from one million information requests in Mexico”, (Berliner, Bagozzi y Palmer-Rubin, 2018).



**Gráfica 5.** Solicitudes a través del tiempo para diez temas de información usando el método de *unsupervised topic modeling*.



**Gráfica 6.** Solicitudes a través del tiempo para seis tipos de información pública creados mediante codificación a mano y *machine learning*.



También categorizamos las solicitudes según el tipo de información solicitada, como una dimensión transversal distinta de los temas de política pública. Dichos tipos de información incluyen “actividades de la dependencia”, “presupuestos y gastos”, “evaluaciones”, “contratos externos”, “personal” e “información regulatoria”. Para el procedimiento de codificar a mano casi 5,000 solicitudes, les pedimos a los expertos en codificación que categorizaran cada solicitud en alguna de las tipos mencionadas o en la categoría de “otro”. Posteriormente, usamos nuestros métodos de *machine learning* para predecir las tipos con base en las palabras o frases que aparecen en cada solicitud, para luego generar puntajes estimados por temática para todas las solicitudes de nuestra base de datos.<sup>5</sup>

La Gráfica 6 desglosa las solicitudes en distintas categorías, pero esta vez, las diferencia por estos tipos de información. En este caso, hay dos temáticas que resaltan en particular a lo largo de todo el periodo de tiempo estudiado: “actividades” y “servidores públicos”. Juntas, estos dos tipos representan alrededor del 70% de las solicitudes cada año. “Contratos” y “presupuestos” representan alrededor del 10% de cada año de manera consistente. Por su parte, “evaluaciones” e “información regulatoria” son considerablemente menores.

Una de las utilidades de nuestros datos es entender *quién* hace las solicitudes de información. En particular, nos interesa determinar qué segmentos de la sociedad mexicana son capaces de usar el sistema. El sistema de acceso a la información de México se diseñó para que fuera de fácil manejo para los usuarios, tanto que un ciudadano común pudiera tener acceso a información sobre el desempeño del gobierno. ¿Ha funcionado así?

Para responder esa pregunta, diseñamos una métrica que nos permite evaluar el grado de expertise que se refleja en cada solicitud de información. La medida es un compuesto de varios elementos del texto de la solicitud, como lenguaje legal, lenguaje técnico y menciones a documentos, fechas, instituciones, organizaciones, personas o lugares específicos. Nuestro equipo de expertos en codificación evaluó cada una de las casi 5,000 solicitudes de entrenamiento para cada una de los nueve variables mencionados; luego, usamos *machine learning* para predecir las temáticas para todas las demás solicitudes con el método de “multi-label prediction” ya descrito. Combinamos las predicciones en una sola medida con el método de análisis de componentes principales. Las solicitudes de información con altos *calificación de expertise* muy probablemente provengan de solicitantes con altos niveles de recursos, formación, experiencia y/o conocimiento interno del funcionamiento de los organismos gubernamentales.

---

<sup>5</sup> Las predicciones se generaron con un método llamado de “multi-label prediction”; más información al respecto en nuestro artículo de 2021 titulado “Multi-Label Prediction for Political Text-as-Data”, (Erlich *et al.*, 2021).

La Gráfica 7 muestra que el promedio de *calificación de expertise* para las solicitudes de información ha aumentado durante el tiempo que lleva de existencia del sistema de acceso a la información en México. Actualmente es más probable que las solicitudes de información presenten mayores señales de expertise que durante los primeros años del programa. La solicitud promedio del último año para el que tenemos datos, 2019, mostró más expertise que el 76.9% de todas las solicitudes de 2003.

El aumento en expertise muy probablemente se deba a dos tipos de cambios: 1) *la composición de los solicitantes*, es decir, que los expertos ahora constituyen un porcentaje mayor del total de los solicitantes y que presentan una mayor cantidad de solicitudes que los no expertos, mucho de los cuales abandonaron por completo el sistema; 2) *aprendizaje por los solicitantes*, significa que algunos de los solicitantes han adoptado dichos marcadores de expertise como estrategia para aumentar la probabilidad de recibir una respuesta satisfactoria. La noción del cambio en la composición tiene implicaciones que podrían afectar la transparencia y el sistema de rendición de cuentas en México, lo que indicaría que dichas instituciones están cada vez menos al alcance del ciudadano común. Por otro lado, sería una señal positiva que el aumento en las calificaciones de expertise se debiera principalmente al aprendizaje, pues reflejaría que la capacidad de los ciudadanos y de la sociedad civil para usar de manera productiva el sistema de acceso a la información está madurando.



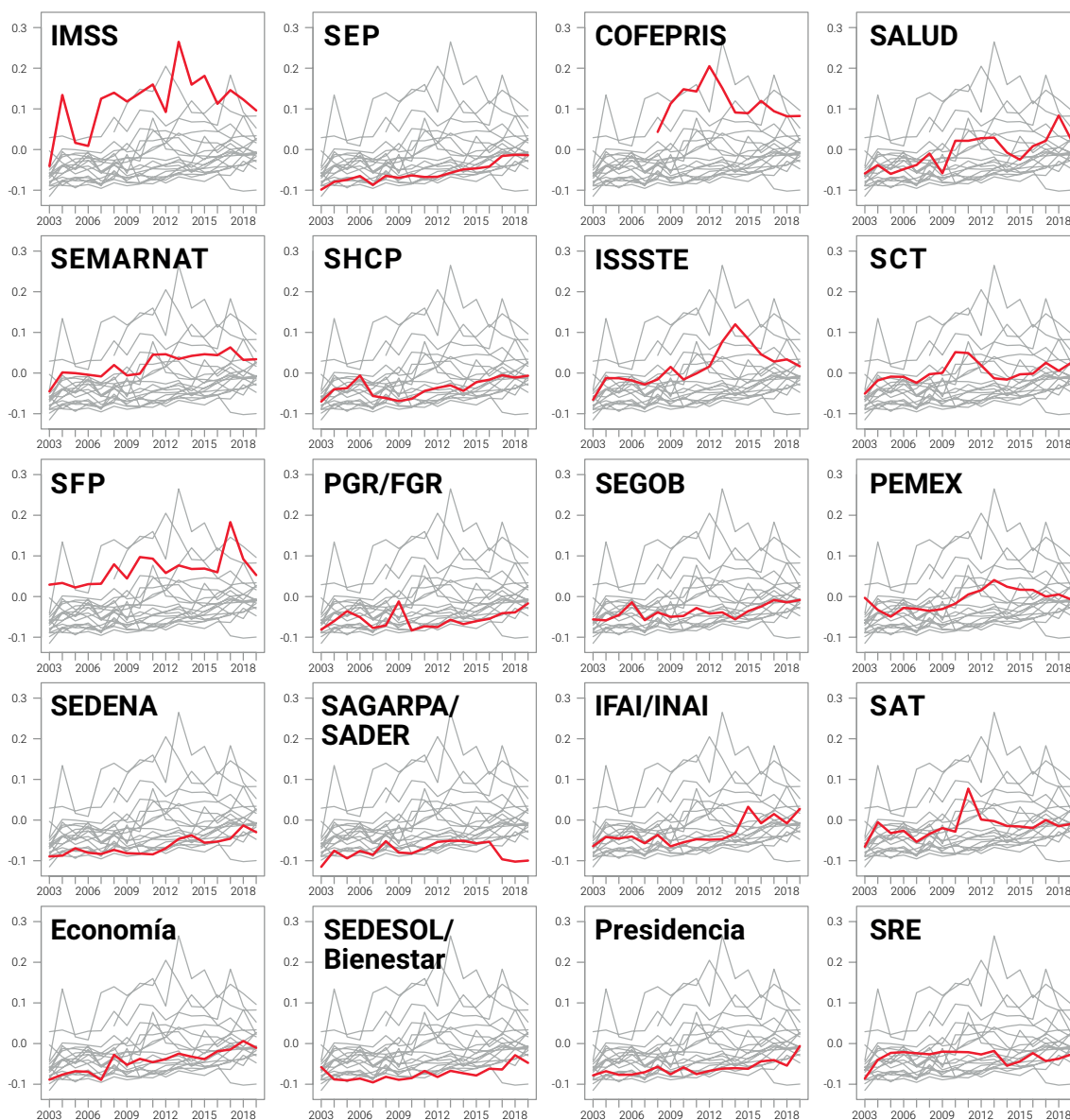
**Gráfica 7.** Calificación de expertise promedio de las solicitudes a través del tiempo.

La Gráfica 8 desglosa la calificación de expertise promedio por estado. Morelos tiene el mayor nivel de expertise promedio; esto se atribuye principalmente al gran número de solicitudes que se describen a sí mismas como parte de una campaña anticorrupción. Lo siguen Baja California y Baja California Sur, cuyas altas calificaciones de expertise promedio podrían estar reflejando las bases de usuarios relativamente limitadas de sus estados.



**Gráfica 8.** Mapa de calificación de expertise promedio de las solicitudes por estado. Un tono más oscuro indica más solicitudes de expertos.

Además, calculamos la calificación de expertise promedio por dependencia, lo que nos permite observar cuáles organismos tienden a recibir solicitudes que demuestran altos niveles de expertise y cuáles tienden a recibir solicitudes con bajos niveles de expertise. Como se muestra en la Gráfica 9, hay una variación significativa, incluso entre las dependencias que reciben una gran cantidad de solicitudes. El IMSS, la dependencia que más solicitudes recibe, muestra de manera consistente un alto nivel de expertise por parte de los solicitantes, mientras que la SEP, la segunda dependencia con mayor cantidad de solicitudes, está en los últimos lugares en términos de la calificación de expertise.



**Gráfica 9.** Calificación de expertise promedio de las solicitudes a través del tiempo para 20 dependencias seleccionadas. Las gráficas individuales de las dependencias están en orden descendente de cantidad de solicitudes recibidas.





# Entender las respuestas

Ahora pasamos a las respuestas de las solicitudes de información. Evaluamos y desagregamos patrones en la capacidad de respuesta de las dependencias. Podemos ver las respuestas de dos formas distintas. De forma concreta, se trata de saber si los ciudadanos realmente obtuvieron una respuesta para la información que solicitaban. Para que una solicitud de información sea exitosa, la información debe presentarse de manera oportuna, corresponder de manera adecuada a la necesidad del solicitante y ser relativamente fácil de entender.<sup>6</sup> De forma un poco más teórica, podemos usar las solicitudes de información como indicadores más amplios de la capacidad de respuesta del gobierno. Una democracia funcional requiere que los ciudadanos puedan pedir cuentas a sus políticos.<sup>7</sup> Al ofrecer información sobre su actividad política, las dependencias de gobierno les dan a los ciudadanos una herramienta necesaria para evaluar, elegir a los políticos y participar en la formulación de políticas públicas. Sin embargo, con mucha frecuencia, políticos y burócratas usan su estatus para ejercer poder sobre los ciudadanos, en lugar de lo contrario. Así, nuestra evaluación sobre la capacidad de respuesta pretende detectar si el sistema de acceso a la información en México es una herramienta efectiva para impulsar la rendición de cuentas de abajo hacia arriba.

Usamos métodos de Big Data de análisis de textos y *machine learning* para construir métricas que midan con mayor precisión la calidad de las respuestas a las solicitudes de información.<sup>8</sup> Además, esas técnicas nos permiten medir la calidad de las respuestas de las más de dos millones de solicitudes de información presentadas ante sujetos obligados federales, sin necesidad de que alguien lea cada una de ellas. En cambio, combinamos la codificación de nuestros expertos de casi 5,000 solicitudes con herramientas de aprendizaje supervisadas para predecir medidas de capacidad de respuesta para el resto. Nuestro equipo de expertos codificó las respuestas de esas solicitudes de entrenamiento para varios elementos clave, por ejemplo, si la respuesta en realidad brindó información o no, si apeló a alguna razón específica para no revelar información (como declararla inexistente o reservada) y si la respuesta ofreció la mitad

---

6 Estas características — “celeridad”, “completitud”, y claridad” — son las tres dimensiones para evaluar respuestas en la Métrica de Gobierno Abierto: [micrositios.inai.org.mx/gobiernoabierto/?page\\_id=5765](https://micrositios.inai.org.mx/gobiernoabierto/?page_id=5765)

7 Ver Schedler (1999).

8 Explicamos a más detalle la lógica de las respuestas en Berliner et al (2021) and Erlich et al (2021).

o más de la información solicitada. Posteriormente, mezclamos esa información con los textos digitalizados de los archivos de respuesta y entrenamos un algoritmo de *machine learning* para predecir los variables que se habían codificado a mano.<sup>9</sup>

De inicio, presentamos los datos que usan medidas de respuestas “ya disponibles”, es decir, que no requieren procesamiento con métodos de Big Data. La Gráfica 10 muestra el cambio en el tiempo de tres medidas importantes tomando en cuenta todas las solicitudes presentadas ante sujetos obligados federales (dentro de México) en la historia del sistema. El primer panel muestra el promedio de días hábiles que tardan las dependencias en responder a las solicitudes de información. La ley federal de acceso a la información estipula un límite inicial de 20 días hábiles para las dependencias, pero podemos considerar que las respuestas que llegan más rápido son mejores, pues es más probable que la información le sea útil al solicitante. Por lo tanto, para esa medida, las cifras más bajas reflejan una mejor capacidad de respuesta. El segundo panel muestra la proporción de solicitudes de información que recibieron una respuesta positiva de “entrega”, donde la dependencia reportó haber brindado la información solicitada.<sup>10</sup> En este caso, un valor más alto refleja una mayor capacidad de respuesta.

El tercer panel identifica la proporción de solicitudes que reciben un tipo particularmente frecuente de respuesta negativa: “inexistencia de la información solicitada”. Expertos de la sociedad civil y académicos han identificado que los servidores públicos muchas veces abusan de esa categoría de respuesta, pues es difícil cuestionarla dado que es imposible demostrar que la información solicitada sí existe.<sup>11</sup> Así, aunque una proporción de las solicitudes obtengan dicha respuesta de manera legítima, seguimos creyendo que sirve como indicador útil para medir el grado en el que el personal de las dependencias evita responder a las solicitudes de información. Para ese indicador, los valores inferiores —un uso menos frecuente de esa categoría de respuesta— señalan mayor capacidad de respuesta.

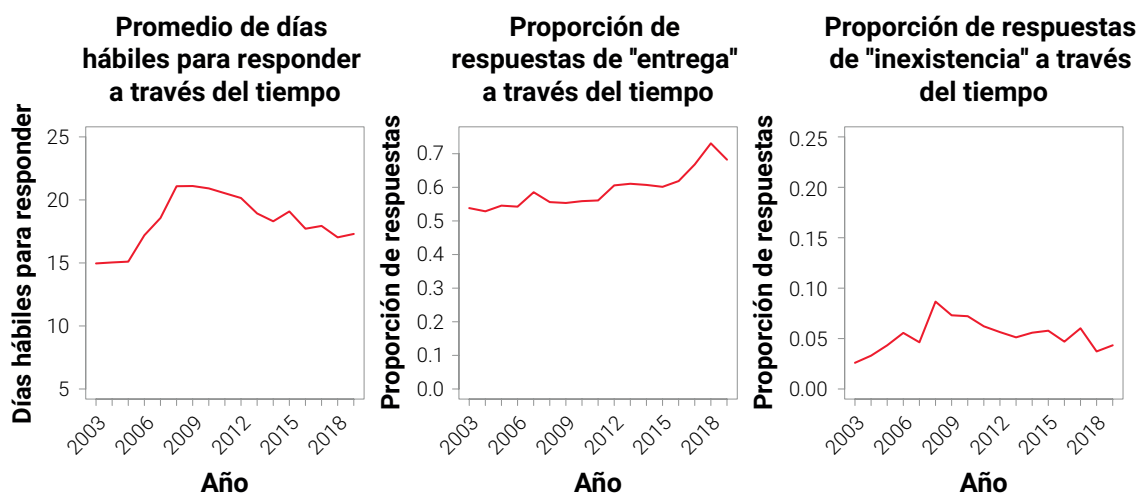
---

9 Esta predicción de *machine learning* usó un método llamado XGBoost, un algoritmo basado en árbol de decisiones. A pesar de la dificultad de predecir los significados de las respuestas con base en los textos, las predicciones funcionaron bastante bien al compararlas con el juicio personal de quienes codificaron a mano. La predicción sobre si se entregó o no cualquier información alcanzó el 79% de precisión; la predicción sobre si una información se declaró inexistente alcanzó el 79.7% de precisión; la predicción sobre si una información se declaró reservada alcanzó el 96.8% de precisión, y la predicción sobre si se reveló la mitad o más de la información solicitada alcanzó el 75.1% de precisión.

10 Clasificamos como “entrega” las respuestas que reportaron mostrar la información en línea (“entrega de información en medio electrónico”), excluyendo el 3.08% de las solicitudes anuales que reportan haber brindado la información en físico. Hay cinco categorías negativas de respuesta: “no corresponde al marco de la ley”, “no es de competencia de la Unidad de Enlace”, “negativa por ser reservada o confidencial”, “inexistencia de la información solicitada” y “no se dará trámite a la solicitud”. Para una definición de estas categorías, véase: [home.inai.org.mx/wp-content/uploads/Definiciones-de-las-variables-estad%C3%ADsticas-2021.pdf](http://home.inai.org.mx/wp-content/uploads/Definiciones-de-las-variables-estad%C3%ADsticas-2021.pdf)

11 Por ejemplo, ver Almánzar *et al.*; (2018) y Fox *et al.*; (2011).

Al analizar esos indicadores de manera conjunta, resulta que la capacidad de respuesta se movió en una dirección negativa durante la primera media década del sistema. Los días promedio para obtener respuesta aumentaron de 15, en 2003, a más de 20 en 2009, mientras que la proporción de solicitudes que recibieron negativas por inexistencia aumentaron de alrededor del 3% a alrededor del 9%. Sin embargo, desde 2009, esos tres indicadores previamente disponibles han mejorado constantemente.



**Gráfica 10.** Indicadores oficiales de capacidad de respuesta usando datos disponibles al público.

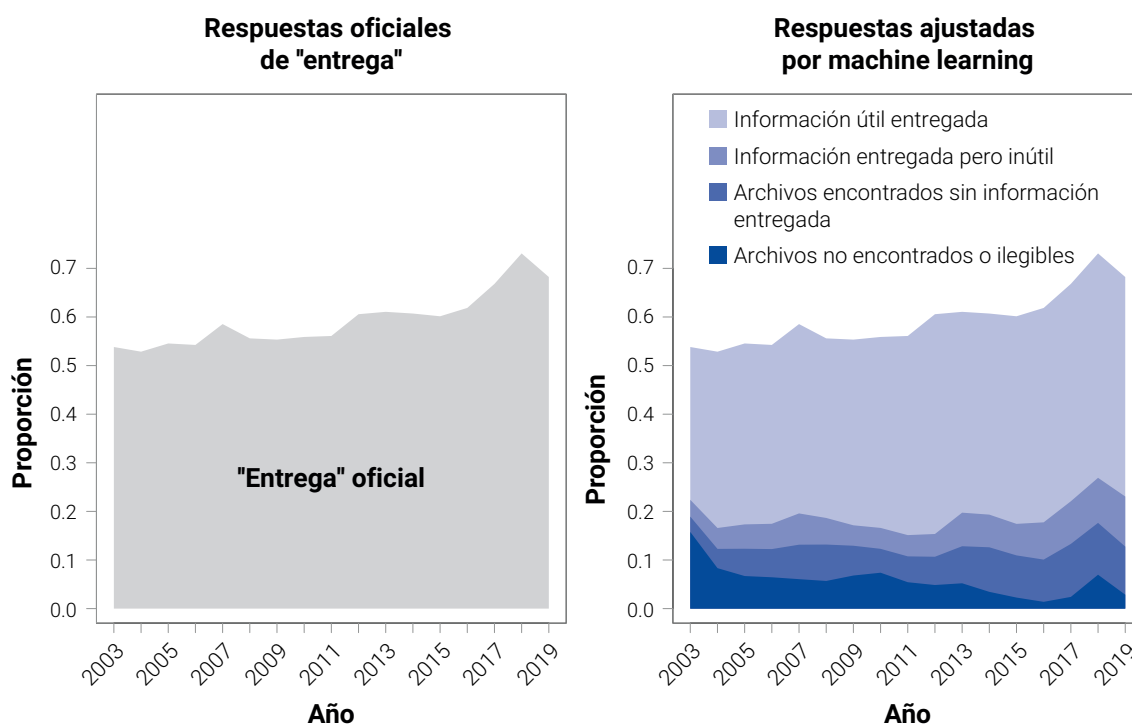
Sin embargo, a partir de entrevistas con miembros de la comunidad mexicana de transparencia y rendición de cuentas, sabemos que las designaciones de las categorías de respuesta muchas veces son erróneas. Por ejemplo, el personal de las dependencias muchas veces reporta “entrega” cuando la respuesta en realidad no brindó ninguna información. También hay un abuso de respuestas negativas clasificadas como “inexistencia”, con las que el personal busca justificar la confidencialidad de la información solicitada incluso aunque esa información debería estar sujeta a requisitos de transparencia.

Nuestras nuevas medidas para la capacidad de respuesta usan *machine learning* para evaluar con mayor precisión el grado en el que las respuestas efectivamente ofrecen la información que el ciudadano buscaba. Es importante resaltar que el algoritmo se basa en información de casi 5,000 pares de solicitud-respuesta que los expertos en gobierno mexicano y en la normatividad en transparencia codificaron “a mano”. Usamos este método para internarnos en la “caja negra” de la información de las respuestas que se etiquetaron como “entrega” para conocer la información que realmente se brindó.

La Gráfica 11 representa un desagregado de todas las respuestas etiquetadas como “entrega” a través del tiempo. El panel de la izquierda —sombreado en gris— simplemente muestra la proporción de solicitudes que recibieron esa etiqueta cada año. La proporción aumenta de alrededor de la mitad de las solicitudes en 2003 a alrededor de dos tercios de las solicitudes de 2019. El panel de la derecha desglosa esa información en cuatro categorías:

- 1 *Información útil entregada*: se brindó la mitad o más de la información solicitada.
- 2 *Información entregada pero inútil*: se brindó información, pero menos de la mitad de lo que se había solicitado.
- 3 *Archivos de respuesta encontrados, pero sin información proporcionada*: no se brindó ninguna información.
- 4 *Archivos de respuesta no encontrados o ilegibles*: imposible hacer predicciones con *machine learning* porque o no servía el enlace Infomex/SNT, o no se encontraron archivos en el enlace o los archivos eran ilegibles.

Podríamos considerar que sólo la primera de esas categorías es una respuesta satisfactoria, pues, en ese caso, se ofreció la mayor parte de la información solicitada. Las otras tres categorías representan alrededor del 20% de todas las solicitudes de información; aunque la proporción aumenta considerablemente desde 2016. Dicho de otro modo, alrededor de un tercio de las solicitudes de información que reciben la respuesta oficial de “entrega” cae en una de las tres categorías insatisfactorias.



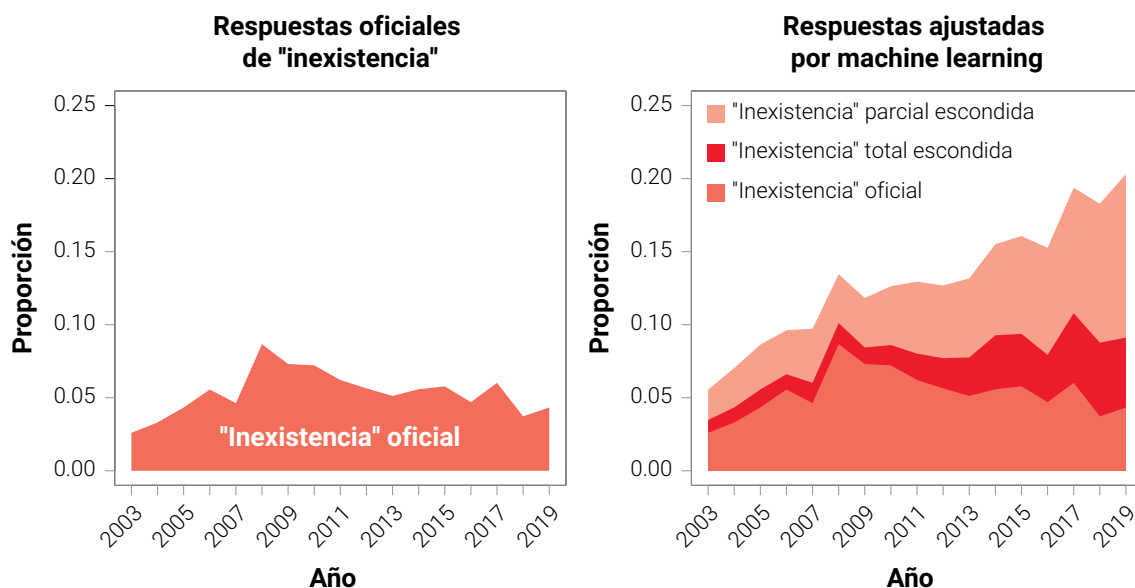
**Gráfica 11.** Comparación de respuestas oficiales de “entrega” a través del tiempo con respuestas reales identificadas mediante predicciones con *machine learning*.

Otro uso de nuestro método de *machine learning* es crear una medida corregida de la frecuencia con la que se niegan las solicitudes de información apelando a la “inexistencia” de la información solicitada. Como se muestra en el panel izquierdo de la Gráfica 12, esta categoría de respuestas alcanzó su pico en 2008 para sumar alrededor del 9% de las solicitudes de información, y desde entonces ha ido en declive, hasta llegar a alrededor del 4% de las respuestas en 2019. Creamos una medida ajustada de “inexistencia” que agrega las respuestas oficialmente clasificadas como “entrega”, pero que, dentro del material de respuesta entregado, en realidad afirman que una parte o toda la información solicitada es inexistente. El panel de la derecha de la Gráfica 12 muestra la frecuencia a través del tiempo de tales respuestas como “inexistencia escondida”. El segmento más oscuro de la gráfica representa las respuestas que oficialmente se designan como “entrega”, pero que la única información que se le brindó al solicitante es una carta afirmando que toda la información solicitada es inexistente. El segmento inmediatamente superior representa las respuestas que brindan parte de la información solicitada, pero que también afirman que hay una porción de información que es inexistente. Si juntamos todas esas categorías, las respuestas de “inexistencia” han aumentado constantemente y, en 2019, alcanzaron más del 20% de las solicitudes. Si sólo consideramos las inexistencias “totales” —las que no ofrecen nada de la información solicitada— esas respuestas se mantuvieron relativamente estables por debajo del 10% entre 2008 y 2019, y, al contrario de lo que aparece en los registros oficiales, no ha habido mejoría.

El fenómeno de las inexistencias escondidas representa una seria amenaza a la legitimidad del sistema mexicano de acceso a la información. Los comisionados del INAI ratificaron esta práctica en una decisión de 2010 como respuesta a un recurso de revisión.<sup>12</sup> Al reconocer la legalidad formal de la práctica, las entrevistas con los actores de la sociedad civil revelaron que esa negación de información ha tenido un efecto desmotivador por dos razones. Primero, por lo general, se abusa de la respuesta de “inexistencia” para información que se sabe debería existir y estar disponible para el público. Sin embargo, es muy difícil cuestionar con éxito esa categoría de negación de información. Un recurso de revisión exitoso podría hacer que el INAI le ordenara a la dependencia realizar una búsqueda exhaustiva de la información, pero si esa búsqueda no lograra presentar la información solicitada, el ciudadano tendría poco que hacer al respecto. En segundo lugar, cuando las negativas por inexistencia se disfrazan de “entregas”, le dificulta al INAI o a los observadores externos evaluar el grado en el que la dependencia realmente está ofreciendo la información solicitada.

---

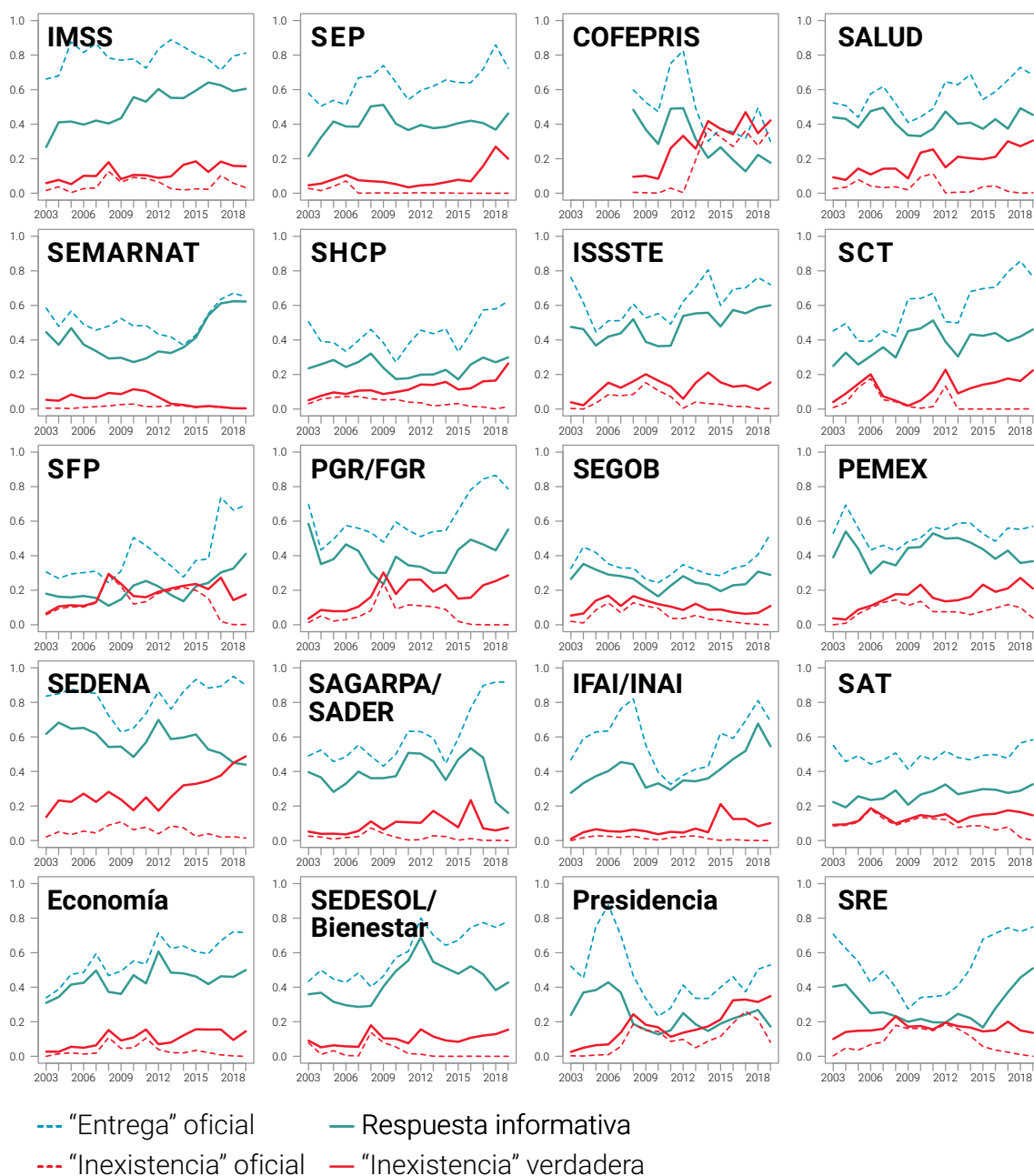
12 Criterio 07/2010: “No será necesario que el Comité de Información declare formalmente la inexistencia cuando del análisis a la normatividad aplicable no se desprenda obligación alguna de contar con la información solicitada ni se advierta algún otro elemento de convicción que apunte a su existencia”. Este principio se reafirmó en Criterio 07/17.



**Gráfica 12.** Comparación de respuestas oficiales de “inexistencia” con respuestas reales identificadas mediante predicciones con *machine learning*. (Nótese que tiene una escala distinta para el eje Y que la de la Gráfica 11).

La Gráfica 13 muestra las mediciones ajustadas de respuesta de 20 dependencias seleccionadas a través del tiempo. Cada panel incluye cuatro líneas. Las líneas azules gruesa y delgada muestran la proporción de respuestas oficialmente designadas como “entrega” y la proporción que nuestros métodos juzgaron que en efecto fueron informativas (es decir, ofrecen al menos la mitad de la información solicitada). La brecha entre esas dos líneas representa la proporción de respuestas clasificadas como “entrega”, pero que no ofrecen información sustancial. Esa brecha ha aumentado en muchas dependencias en los últimos años. Sobresale la SADER: en 2019, clasificó como “entrega” el 91.6% de las respuestas, pero estimamos que sólo el 16% de las respuestas fueron informativas. En el otro extremo, la SEMARNAT sobresale como una dependencia con muy baja incidencia de dar respuestas poco informativas. Desde 2014, prácticamente todas las respuestas que la SEMARNAT reporta como “entrega” ofrecieron al menos la mitad de la información solicitada. Este ajuste arroja luz nueva sobre el nivel de cumplimiento de la SEMARNAT con los requisitos de transparencia. Aunque su incidencia oficial de respuestas marcadas como “entrega” esté en la parte baja, la medida corregida de respuestas informativas sobresale del promedio.

Las dos líneas rojas para cada dependencia en la Gráfica 13 reflejan la proporción de respuestas oficialmente designadas como “inexistencia” (línea delgada) y la proporción aumentada que incluye negativas totales o parciales disfrazadas de “inexistencia” a partir de nuestro método de *machine learning* (línea gruesa). Para muchas dependencias, el uso de la etiqueta “inexistencia” de manera escondida ha aumentado en los últimos años. Por ejemplo, la Secretaría de Salud ha reportado “inexistencia” para menos del 5% de las solicitudes cada año desde 2012, mientras que nuestra medida de “inexistencia” real revela que más del 20% —que rebasó 30% en 2017 y 2019— de las respuestas niegan la información. La SHCP, el ISSSTE, la SCT, la PGR/FGR y la SEDENA presentan tendencias similares: muy pocas “inexistencias” reportadas, pero proporciones crecientes de “inexistencia” real.



**Gráfica 13.** Comparación de respuestas oficiales y reales a través del tiempo para 20 dependencias seleccionadas. Las gráficas individuales de las dependencias están en orden descendente de solicitudes recibidas. Las líneas delgadas muestran las respuestas oficiales de "entrega" e "inexistencia". Las líneas gruesas muestran respuestas informativas (aquellas para las que se entregó la mitad o más de la información solicitada) y las respuestas de "inexistencia" real (ya sea total o parcial).

Las tablas de esta sección identifican el mayor y menor puntaje de los sujetos obligados durante los últimos cinco años de los datos (2015-19) para las dos mediciones clave que creamos con nuestro método de *machine learning*. La Tabla 1 muestra las dependencias que con mayor frecuencia ofrecieron una respuesta informativa a las solicitudes, y las diez dependencias que lo hicieron con la menor frecuencia (a partir de los 50 sujetos obligados con la mayor cantidad de solicitudes recibidas durante el periodo). La PROFECO aparece como la dependencia que con mayor frecuencia ofreció una respuesta informativa, y la Comisión Nacional Bancaria aparece como la dependencia que lo hizo con menor frecuencia.



Sujeto Obligado (primeros y últimos diez lugares)	% Respuesta informativa
Procuraduría Federal del Consumidor	67.6
Instituto Politécnico Nacional	66.1
Instituto Nacional de Migración	65.1
Universidad Nacional Autónoma de México	62.2
Instituto Mexicano del Seguro Social	61.1
Consejo Nacional de Ciencia y Tecnología	61.0
Aeropuertos y Servicios Auxiliares	59.4
Instituto Nacional de Antropología e Historia	59.2
Comisión Nacional de Áreas Naturales Protegidas	57.9
Secretaría de Medio Ambiente y Recursos Naturales	57.1
...	...
Servicio de Administración Tributaria	29.8
Secretariado Ejecutivo del Sistema Nacional de Seguridad Pública	27.3
Archivo General de la Nación	27.1
Secretaría de Hacienda y Crédito Público	26.5
Registro Agrario Nacional	25.3
Secretaría de Gobernación	24.9
Presidencia de la República	20.4
Pemex Transformación Industrial	19.8
Comisión Federal para la Protección contra Riesgos Sanitarios	18.3
Comisión Nacional Bancaria y de Valores	17.0

**Tabla 1.** Primeros y últimos diez lugares de los sujetos obligados, por porcentajes de respuestas informativas, 2015-19. (De los 50 sujetos obligados con la mayor cantidad de solicitudes durante el periodo).

La Tabla 2 usa nuestra medida corregida de respuestas de “inexistencia” para crear un nuevo puntaje de las dependencias que usaron esa categoría con mayor y menor frecuencia de entre las 50 dependencias que más solicitudes recibieron. Como se muestra en la Gráfica 11, la SEMARNAT pocas veces niega información a través de la categoría de “inexistencia” (ni de manera oficial ni encubierta). En total, sólo ocho dependencias usan esa categoría de negación para menos del 10% de las solicitudes. Por otro lado, las diez dependencias que con mayor frecuencia usan la etiqueta de “inexistencia” lo hicieron más del 25% de las veces. Sobresalen tres dependencias que afirman inexistencia más del 40% de las veces: la SEDENA, la COFEPRIS y la Comisión Nacional Bancaria.

Sujeto Obligado (primeros y últimos diez lugares)	% Inexistencia real
Secretaría de Medio Ambiente y Recursos Naturales	1.0
Instituto Mexicano de la Propiedad Industrial	3.9
Consejo Nacional de Ciencia y Tecnología	6.2
Secretaría de Gobernación	8.3
Aeropuertos y Servicios Auxiliares	9.0
Archivo General de la Nación	9.3
Comisión Federal de Electricidad	9.8
Secretaría de Agricultura Ganadería Desarrollo Rural Pesca y Alimentación/Secretaría de Agricultura y Desarrollo Rural	10.1
Instituto del Fondo Nacional de la Vivienda para los Trabajadores	10.1
Secretaría Del Trabajo y Previsión Social	10.2
...	...
Comisión Nacional de Áreas Naturales Protegidas	25.2
Instituto Nacional de Antropología e Historia	25.4
Secretaría de Salud	26
Policía Federal	27.7
Lotería Nacional para la Asistencia Pública	28.5
Presidencia de la República	31.2
Secretaría de Marina	33.4
Comisión Federal para la Protección contra Riesgos Sanitarios	40.1
Comisión Nacional Bancaria y de Valores	40.1
Secretaría de la Defensa Nacional	40.8

**Tabla 2.** Primeros y últimos diez lugares de los sujetos obligados, por porcentajes de respuestas de inexistencia real, 2015-19. (De los 50 sujetos obligados con la mayor cantidad de solicitudes durante el periodo).



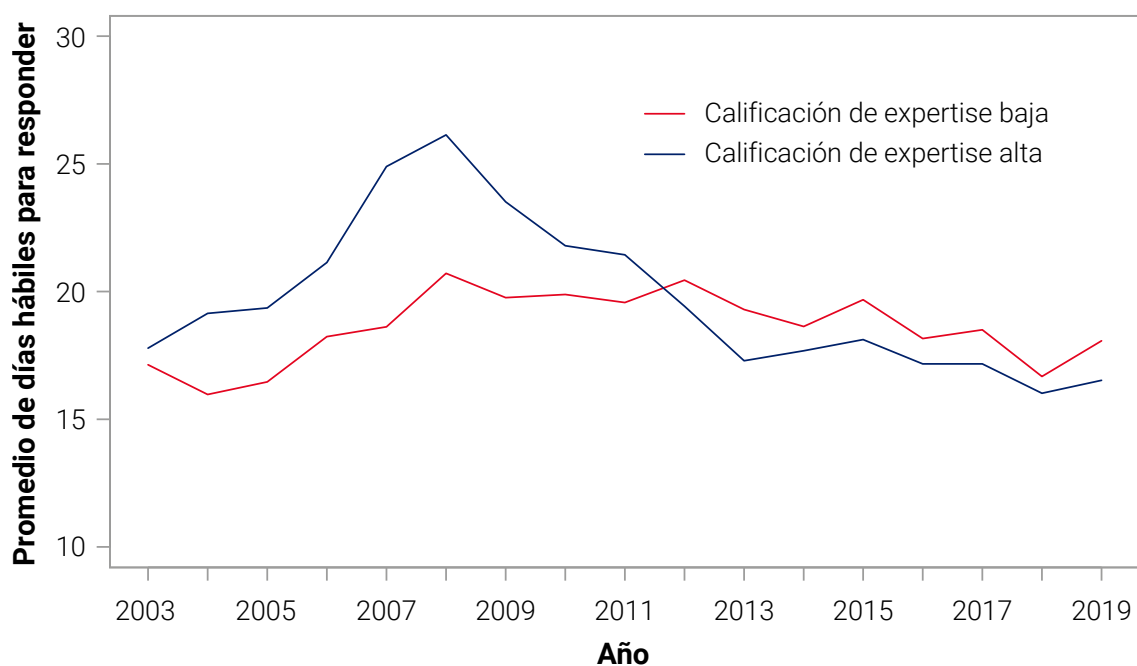
# Sesgo de Expertise

Concluimos nuestro análisis cruzando la calificación de expertise para las solicitudes con las medidas de capacidad de respuesta. Cruzar esas dos variables estimadas con *machine learning* nos permite detectar los sesgos de los expertos (o de los no expertos) en la capacidad de respuesta. Es decir, nos interesa observar las diferencias en la capacidad de respuesta para las solicitudes de información presentadas por solicitantes expertos versus no expertos. Las Gráficas 14 y 15 contienen cada una dos líneas que grafican las medidas de capacidad de respuesta a través del tiempo, promedio de días hábiles para responder y proporción de respuestas que reciben una “verdadera” (corregida con *machine learning*) respuesta de inexistencia. En ambas gráficas, la línea morada representa la capacidad de respuesta a las solicitudes de información que se encuentran en el cuartil superior de calificación de expertise, y la línea naranja representa la capacidad de respuesta a las solicitudes que están en el cuartil inferior. En general, el sistema inició con un sesgo hacia los solicitantes no expertos, pero con el tiempo evolucionó para reaccionar mejor a los solicitantes expertos.<sup>13</sup>

Por ejemplo, la Gráfica 14 muestra los días promedio para responder a solicitudes con la calificación de expertise más altos versus los más bajos manteniendo constante el formato de la solicitud, la complejidad y la temática. Entre 2003 y 2011, las solicitudes con altas calificaciones de expertise se tardaron más en recibir respuestas que aquellas con calificaciones de expertise más bajas. Esta brecha alcanzó su máximo en 2008, cuando el tiempo para dar respuesta a las solicitudes de expertos era de cinco días hábiles más (un promedio de alrededor de 26 días hábiles, comparado con 20 días hábiles para las solicitudes de bajo expertise). Sin embargo, a partir de 2012 observamos respuestas ligeramente más rápidas para las solicitudes que demuestran mucho expertise.

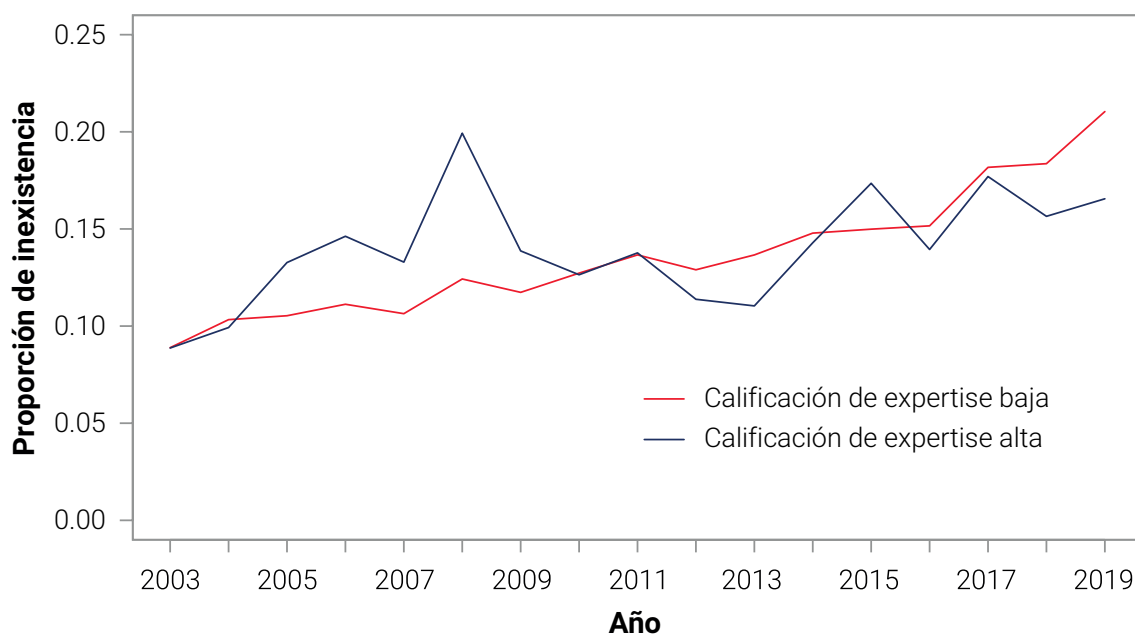
---

13 Como las solicitudes de expertos muchas veces presentan temas más complejos, estas medidas se ajustaron para controlar por diferencias en el formato, la complejidad y la temática de cada solicitud.



**Gráfica 14.** Días promedio para recibir respuesta para solicitudes de información en los cuartiles de expertise superior e inferior (ajustados para controlar por diferencias en el formato, la complejidad y la temática de cada solicitud).

La Gráfica 15 muestra resultados similares con una medida distinta de capacidad de respuesta: la incidencia de negaciones por inexistencia. Una vez más, vemos un mayor nivel de capacidad de respuesta a las solicitudes de no expertos durante los primeros años del sistema —manteniendo todo lo demás constante— sobre todo entre 2005 y 2009, cuando la proporción de solicitudes de alto expertise que recibieron negaciones por inexistencia fue al menos 2 puntos porcentuales superior a la proporción de solicitudes de bajo expertise que recibieron ese tipo de respuesta. Sin embargo, a partir de 2010, las negativas por inexistencia se han presentado en la misma magnitud para las solicitudes de bajo expertise que para las de alta, y en algunos años, incluso en magnitud superior.



**Gráfica 15.** Tasa promedio de “inexistencia verdadera” para solicitudes de información en los cuartiles de expertise superior e inferior (ajustada para controlar por diferencias en el formato, la complejidad y la temática de cada solicitud).

¿Qué podría explicar esta transición en la capacidad de respuesta? Hay dos explicaciones posibles para las ganancias relativas de una buena capacidad de respuesta para las solicitudes de información con altas calificaciones de expertise frente a las solicitudes con calificaciones bajas. Primero, podría ser simplemente que los expertos han aprendido a presentar mejor las solicitudes de información, por ejemplo, a ser más precisos al describir la información que buscan o a haber desarrollado una mejor comprensión de qué tipo de información realmente está disponible al público que los no expertos. Segundo, podría deberse a que el personal de las dependencias ha aprendido ya sea: a) a desarrollar mejores sistemas para manejar tipos de solicitudes más complejos que muy probablemente sean presentadas por solicitantes expertos; o b) a dar prioridad a las solicitudes que muestren mayor expertise, quizá porque estén conscientes de que las solicitudes de expertos son más proclives a presentar recursos de revisión ante respuestas insatisfactorias. Con base en una investigación cualitativa con solicitantes de información de la sociedad civil y con personal de las Unidades de Transparencia, suponemos que dicha transformación refleja los procesos que ocurren tanto entre los ciudadanos solicitantes como entre el personal de las dependencias de gobierno.



# Conclusiones y recomendaciones

Los hallazgos de este proyecto dan motivo de celebración y de preocupación al mismo tiempo tras el análisis del sistema de acceso a la información en México y su potencial para contribuir con los procesos democráticos de la rendición de cuentas. Por el lado positivo, el sistema claramente se ha consolidado y se ha convertido en una herramienta central para una amplia variedad de actores clave en la democracia en México: organizaciones de la sociedad civil, periodistas y el sector privado, para mencionar algunos. Con métodos de *Big Data* y *Data Science* que nos permitieron observar el cambio en el lenguaje que se usa en las solicitudes de información a través del tiempo, detectamos un nivel creciente de expertise en la solicitud promedio, lo que sugiere que los actores están desarrollando capacidades para acceder con mayor éxito a la burocracia que proporciona información. Además, las dependencias del gobierno federal han acumulado recursos y competencias para responder de manera puntual a miles de solicitudes de información de manera semanal. Aunque este estudio no haya tomado en cuenta solicitudes de información presentadas ante gobiernos municipales ni estatales, éstos también se han insertado en la infraestructura gubernamental para brindar información pública. En suma, el sistema de acceso a la información en México ha superado la promesa que hizo al fundarse hace dos décadas.

Por otro lado, la consolidación del sistema de acceso a la información ha estado acompañada de algunas tendencias que han ido en contra del uso productivo del sistema de transparencia, sobre todo por parte de los ciudadanos no expertos. El personal de gobierno que opera dentro de los sujetos obligados encargados de responder a las solicitudes de información ha desarrollado una serie de tácticas para evitar responder a las solicitudes de información. Una de las más comunes es afirmar que la información solicitada no existe. Aunque ésta sea una justificación legalmente válida para negar información, los actores de la sociedad civil afirman que se usa frecuentemente en casos donde la información debería estar disponible para los ciudadanos solicitantes. Además, muchas de esas negativa por inexistencia están mal etiquetadas como entrega de información, lo que reduce la capacidad de los evaluadores para medir la frecuencia de esa categoría de negación con métodos convencionales. Nuestros métodos de *machine learning* nos permitieron abrir la “caja negra” de las respuestas a solicitudes de información para construir estimados más precisos de la frecuencia con la que en realidad se entregan esas negativa. Un hallazgo clave es que algunas dependencias son mucho más honestas a la hora de etiquetar sus respuestas como negativas, mientras que otras etiquetan de forma errónea las respuestas negativas con bastante regularidad.

Nuestra investigación también subraya el gran potencial del *machine learning* y de otras herramientas de análisis cuantitativo para que alguien más los aplique y proporcione más datos sobre el acceso a la información en México. Tanto el propio INAI como las organizaciones de la sociedad civil podrían extender y actualizar un análisis como el que presentamos aquí. El INAI en particular podría usar estas herramientas junto con su gran corpus de evaluaciones y datos internos para generar medidas, sistemas de monitoreo y recomendaciones más detalladas. Sin embargo, hay algunas barreras técnicas que dificultaron nuestro proyecto y que podrían dificultar también otros proyectos. La más importante de las barreras fue que las respuestas a las solicitudes de información muchas veces se entregan en archivos PDF difíciles de digitalizar debido a la baja calidad del archivo, así como a los sellos y firmas que oscurecen el texto. Dichos archivos no sólo son difíciles de usar para los solicitantes, sino que son barreras serias para generar herramientas de *machine learning* que evalúen de manera adecuada la calidad de la respuesta. Una recomendación sería exigir que, además de que las cartas de respuesta se suban a la Plataforma Nacional de Transparencia en formato PDF, se introdujera el contenido de la carta en un cuadro de texto para que quedara guardado en las bases de datos internas del INAI. Esto permitiría una mejor evaluación y análisis, sin abandonar por completo los documentos en PDF.

Aunque estemos convencidos de la promesa que representan estas herramientas de Big Data para mejorar el sistema de acceso a la información de México, también estamos comprometidos con la noción de que el análisis cuantitativo por sí mismo es insuficiente para contar la historia completa. Nuestros hallazgos estuvieron inspirados y confirmados por cientos de horas de exploración cualitativa en el sistema de acceso a la información en México: 1) leer miles de solicitudes de información y sus respuestas; 2) entrevistar a representantes de docenas de asociaciones de la sociedad civil que solicitan información como parte de sus actividades; 3) entrevistar personal del INAI y las Unidades de Transparencia y 4) leer y analizar periodismo de investigación basado en información reunida a través de solicitudes de información. Y mientras que este proyecto ha contado con un equipo bastante grande de investigadores durante varios años, podemos visualizar fuentes adicionales de datos para enriquecer nuestros hallazgos. Por ejemplo, no logramos recopilar datos detallados sobre recursos y procesos de las Unidades de Transparencia dentro de cada sujeto obligado. Tales datos serían muy útiles para probar algunas de las hipótesis que surgieron a lo largo de la investigación. En suma, una apreciación holística de las instituciones de rendición de cuentas de México —o de cualquier país— requiere un enfoque integral.

En términos generales, las tendencias que descubrimos en este estudio reflejan una profesionalización del ecosistema de transparencia y rendición de cuentas en México. Aunque lo consideramos positivo, sugerimos tres enfoques estratégicos de políticas públicas para adaptar el cambiante panorama ATI:



Primero, con base en ese análisis holístico, el personal del INAI podría implementar un proceso de entrenamiento más exhaustivo y enfocado para el personal de las Unidades de Transparencia de las dependencias para ayudar a diseminar mejores prácticas y para estandarizar métodos. Uno de los hallazgos importantes de nuestro estudio es un alto grado de heterogeneidad en las prácticas de respuesta entre dependencias, que se refleja, por ejemplo, en niveles muy distintos de negativas por inexistencia.

Segundo, defendemos un énfasis renovado para desarrollar sistemas estandarizados de transparencia que les permitan a los ciudadanos acceder a información importante sin necesidad de presentar una solicitud de información. El elevado uso del sistema de solicitudes de información es un logro importante del INAI. Sin embargo, esperamos que en el futuro la cantidad de solicitudes disminuya conforme los ciudadanos vayan encontrando la información que buscan en sitios o publicaciones del gobierno. La divulgación proactiva de tipos de información frecuentemente solicitadas traería dos beneficios: primero, democratizaría y simplificaría el sistema de solicitud de información, contrarrestando el sesgo de expertise y reduciendo el tiempo de espera de la información que buscan los solicitantes; y segundo, reduciría la carga de trabajo del personal encargado de procesar y responder a las solicitudes de información en cada dependencia.

Nuestra tercera recomendación es tomar lo que parece ser una debilidad del sistema de acceso a la información de México —el creciente sesgo de expertise— y convertirlo en una fortaleza. La idea sería formalizar el papel de las OSC como consultores o intermediarios para acceder a la información pública. Quizá el gran ecosistema de organizaciones de la sociedad civil enfocadas en transparencia y rendición de cuentas que ha desarrollado competencias y experiencia para manejar el sistema ha sido tan importante como el INAI (el “instituto garante”) para el funcionamiento del sistema. Dichas organizaciones se han unido en redes como la Red por la Rendición de Cuentas y Alianza para el Gobierno Abierto para involucrarse en la supervisión y la mejoría de las instituciones para la transparencia y el buen gobierno que están consolidándose en México. El INAI podría asociarse con esas organizaciones y otorgarles financiamiento, capacitación y estatus oficial como actores clave a los que pueden recurrir los no expertos para que los ayuden a emitir sus solicitudes de información. La sociedad civil ha tenido un papel clave diseminando y supervisando las instituciones de rendición de cuentas en todo el mundo. El sistema de acceso a la información en México sería buen candidato para construir alianzas productivas con esos actores.



# Referencias

Almanzar, Tanya, Mark Aspinwall y David Crow (2018), "Freedom of Information in Times of Crisis: The Case of Mexico's War on Drugs", *Governance* 31(2): 321–339.

Berliner, Daniel, Benjamin E. Bagozzi y Brian Palmer-Rubin (2018), "What Information Do Citizens Want? Evidence from One Million Information Requests in Mexico", *World Development* 109: 222–235.

Berliner, Daniel, Benjamin E. Bagozzi, Brian Palmer-Rubin y Aaron Erlich. 2021. "The Political Logic of Government Disclosure: Evidence from Information Requests in Mexico." *The Journal of Politics* 83(1):1–18.

Cejudo, Guillermo M., Sergio López Ayllón y Alejandra Ríos Cázares (2014), *La Política de Transparencia en México: Instituciones, Logros y Desafíos*, Ciudad de México: CIDE.

Erlich, Aaron, Daniel Berliner, Brian Palmer-Rubin y Benjamin E. Bagozzi. 2021. "Media Attention and Bureaucratic Responsiveness." *Journal of Public Administration Research and Theory* 31(4):687–703.

Erlich, Aaron, Stefano G. Dantas, Benjamin E. Bagozzi, Daniel Berliner y Brian Palmer-Rubin (2021), "Multi-Label Prediction for Political Text-as-Data", *Political Analysis*: 1–18. DOI: [doi.org/10.1017/pan.2021.15](https://doi.org/10.1017/pan.2021.15)

Fox, Jonathan A., Libby Haight y Brian Palmer-Rubin (2011), "Proporcionar Transparencia: ¿Hasta Qué Punto Responde El Gobierno Mexicano a Las Solicitudes de Información Pública?", *Gestión y Política Pública* 20(1): 3–61.

Fung, Archon, Mary Graham y David Weil (2007), *Full Disclosure: The Perils and Promise of Transparency*, Cambridge University Press.

Lagunes, Paul y Oscar Pocasangre (2019), "Dynamic Transparency: An Audit of Mexico's Freedom of Information Act", *Public Administration* 97(1): 162–176.

Schedler, Andreas, Larry Jay Diamond y Marc F. Plattner (1999), *The Self-Restraining State: Power and Accountability in New Democracies*, Boulder, CO: Lynne Rienner Publishers.

# Equipo de investigación



## Daniel Berliner

Es profesor investigador de Ciencias Políticas y Políticas Públicas en el Departamento de Gobierno de London School of Economics and Political Science. Su investigación principal se centra en el rol de la información en la política y la gobernanza, incluyendo el estudio del impacto de la información en el comportamiento de los votantes y de la burocracia, las políticas de transparencia, la rendición de cuentas, reformas a la participación política, e iniciativas de reforma a la gobernanza impulsadas por actores transnacionales. Utilizando métodos cualitativos y cuantitativos

ha realizado análisis en México, Sudáfrica, Estados Unidos y del contexto global. En 2012, obtuvo el grado de Doctor por la Universidad de Washington en Seattle Estados Unidos. Previamente a su estadía en LSE, mantuvo distintas posiciones en Freie Universität Berlin, University of Minnesota, y Arizona State University. Su investigación ha sido publicada en *the American Political Science Review*, *The Journal of Politics*, *British Journal of Political Science*, y *Journal of Public Administration Research and Theory*. Los detalles de su investigación pueden ser consultados en su página personal: [danielberliner.com/](http://danielberliner.com/)



## Brian Palmer-Rubin

Es profesor investigador en Ciencias Políticas en Marquette University. Obtuvo el grado de doctor en Ciencias Políticas por la Universidad de California, Berkeley y cuenta con una maestría en Estudios Latinoamericanos por la Universidad Nacional Autónoma de México (UNAM). Sus temas de investigación son la representación de intereses, la economía política del desarrollo y las instituciones de acceso a la información con especial atención al caso mexicano. Su libro *Evading the Patronage Trap: Interest Representation in Mexico*, (a publicar en julio de 2022

por la Universidad de Michigan) examina cómo ciertos grupos de interés representan a su sector en la determinación de políticas públicas, y ganó el premio a la mejor tesis en políticas públicas Harold D. Lasswell 2016 de la Asociación Americana de Ciencias Políticas (APSA). Otros trabajos de investigación han sido publicados en revistas como: *Comparative Political Studies*, *Journal of Politics*, *Political Analysis*, *Studies in Comparative International Development*, *World Development*, *Gestión y Política Pública*, y otras. Sus publicaciones y más información sobre el perfil académico de Brian se encuentran en:

[brianpalmerrubin.com](http://brianpalmerrubin.com)



### **Jéssica E. Tapia Reyes**

Cuenta con más de 15 años de experiencia en el análisis y evaluación de políticas públicas. Trabajó para el Gobierno Federal en México, en distintas instituciones, principalmente se desempeñó en la evaluación y diseño de políticas públicas en áreas como comercio exterior, transparencia, ciencia y tecnología, y desarrollo urbano. Fue premio nacional de la administración pública, nivel Secretaría, en la Secretaría de Economía en 2012 y fue delegada Nacional y ponente ante

el “*Domestic Paperless Trading Environment Workshop*” del Foro de Cooperación Económica Asia Pacífico (APEC) en 2010 y 2012. Su formación y práctica profesional se concentran en la Ciencia de Datos y su aplicación a problemas sociales y al análisis de políticas públicas. Actualmente además de coordinar el proyecto Big Data y Acceso a la Información en México, colabora en el desarrollo de diversos proyectos en los que se intersectan el análisis de política pública, el uso masivo de datos y la aplicación de algoritmos de aprendizaje de máquina, en especial el análisis de lenguaje natural y la minería de texto. Forma parte del programa “Skills For Women in Tech” de la Embajada del Reino Unido en México.



### **Benjamin E. Bagozzi**

Es profesor investigador en el Departamento de Ciencias Políticas y Relaciones Internacionales de la Universidad de Delaware. Tanto sus actividades de docencia, como su investigación, se enfocan a las ciencias sociales computacionales, el análisis automático de texto con Big Data e Inteligencia Artificial, derechos humanos, violencia política y política ambiental. Es Doctor en Ciencias Políticas por Penn State University desde 2013. Sus investigaciones han sido

publicadas en *Journal of Politics*, *Political Analysis*, *Journal of Conflict Resolution*, *World Development*, *Journal of Peace Research*, y *Sociological Methods & Research*, entre otros medios. Su sitio web profesional y su laboratorio (Social Data Analytics Lab – SADL) pueden ser consultados en la siguiente página web: [benjaminbagozzi.com/](http://benjaminbagozzi.com/)



### Aaron Erlich

Es profesor investigador en el Departamento de Ciencias Políticas de la Universidad de McGill, afiliado al Centro para la Ciudadanía Democrática, al Centro sobre la Dinámica de la Población y al Centro para la Ciencia de Datos Sociales y Culturales. Mediante el uso de encuesta, así como de técnicas de aprendizaje automático, la investigación del Dr. Erlich aborda el papel que juega la información en las políticas democráticas. Entre sus intereses se encuentra la promoción del uso de

métodos cuantitativos para medir el efecto de la información. A nivel regional, el Dr. Erlich ha conducido diversos estudios de gran escala en Europa del Este, África subsahariana y América del Norte. Los detalles de su investigación pueden ser consultados en su página personal: [aaronerlich.com/](http://aaronerlich.com/)

Berliner, Daniel, Brian Palmer-Rubin, Jéscica E. Tapia Reyes, Benjamin E. Bagozzi, Aaron Erlich. (2022). *Big data y acceso a la información en México* (Informe de la Política Pública). The London School of Economics and Political Science. URL: [bigdataytransparenciamx.lse.ac.uk/](http://bigdataytransparenciamx.lse.ac.uk/)

Supported by the LSE Knowledge Exchange and Impact Fund, the LSE Department of Government, the LSE Suntory and Toyota International Centres for Economics and Related Disciplines, Marquette University Committee on Research, Social Science and Humanities Research Council (SSHRC) (grant no. 430-2018-1069), and Fonds de recherche du Québec—Société et culture (FQRSC) (grant no. 253243).



[lse.ac.uk/government](https://lse.ac.uk/government)



Department of Government  
The London School of Economics  
and Political Science  
Houghton Street  
London WC2A 2AE

Email: [d.berliner@lse.ac.uk](mailto:d.berliner@lse.ac.uk)

The London School of Economics and Political Science is a School of the University of London. It is a charity and is incorporated in England as a company limited by guarantee under the Companies Acts (Reg no 70527).

The School seeks to ensure that people are treated equitably, regardless of age, disability, race, nationality, ethnic or national origin, gender, religion, sexual orientation or personal circumstances.

Design: LSE Design Unit ([info.lse.ac.uk/staff/divisions/communications-division/design-unit](mailto:info.lse.ac.uk/staff/divisions/communications-division/design-unit))



Cover image: [istock.com](https://www.istock.com)