

La evidencia como bien público: una matriz de gobernanza para la rendición de cuentas en políticas públicas

LEONARDO ADALBERTO GATICA ARREOLA
PATRICIA MURRIETA CUMMINGS

Recibido: 31 de octubre de 2025. Aceptado: 1 de marzo de 2026.

Resumen. Este artículo examina en qué condiciones la evidencia puede operar como bien público, generando transparencia y rendición de cuentas. Sostenemos que la evidencia adquiere valor público cuando su gobernanza garantiza apertura, inteligibilidad, participación, integridad y salvaguardas de derechos. Con base en ello, proponemos un modelo de gobernanza de la evidencia que permite el uso de la evidencia como bien público.

Palabras clave: evaluación, bienes públicos, gobernanza, participación, derechos humanos.


Evidence as a public good: a governance matrix for accountability in public policies

Abstract. This article examines the conditions under which evidence can operate as a public good, generating transparency and accountability. We argue that evidence acquires public value when its governance ensures openness, intelligibility, participation, integrity, and safeguards for rights. Based on this premise, we propose a model of evidence governance that enables the use of evidence as a public good.

Keywords: evaluation, public goods, governance, participation, human rights.

.....
Leonardo Adalberto Gatica Arreola. Universidad de Guadalajara, México (correo electrónico: leonardo.gatica@academicos.udg.mx | <https://orcid.org/0000-0002-4387-3608>).

Patricia Murrieta Cummings. Universidad de Guadalajara, México (correo electrónico: patricia.murrieta@academicos.udg.mx | <https://orcid.org/0000-0002-9076-3276>).

DOI: <https://doi.org/10.32870/rc.vi8.177> | Esta obra está bajo una Licencia Creative Commons Atribución-NoComercial-SinDerivadas 4.0 Internacional. 

Introducción

¿Puede la producción de evidencia convertirse en un bien público que alimente la transparencia y participación ciudadana en torno a las políticas sociales dirigidas a las infancias? Aun con los cambios recientes en México y otros países de América Latina, podemos afirmar que la región ha avanzado en transparencia y consolidación de marcos normativos robustos para la rendición de cuentas y el acceso a la información. Sin embargo, como argumentan Fung *et al.* (2007), no basta con esto. Si bien es un elemento esencial, el acceso a la información no garantiza por sí mismo un mayor control democrático ni la participación informada cuando la evidencia es difícil de entender, no considera la opinión de los grupos afectados o carece de autonomía respecto a las instituciones que evalúa. La información no garantiza la transparencia ni la rendición de cuentas (Fung *et al.*, 2007; Fox, 2015).

Graizbord (2024) muestra que la «evidencia» no llega al espacio público como un insumo neutro; se construye mediante trabajo organizacional, negociaciones interinstitucionales y la creación de fronteras entre lo «técnico» y lo «político». En tal sentido, el valor público de la evidencia depende de reglas de apertura y de las condiciones socioinstitucionales que permiten producirla, interpretarla y disputarla en términos públicos.

En este artículo proponemos considerar la producción y el uso de la evidencia como un bien público cuya problematización tiene implicaciones distributivas. Como lo plantea Parkhurst (2017), consideramos que quién decide qué, cómo y para qué se mide afecta la legitimidad de las políticas sociales y la redistribución del poder y los recursos entre el Estado, la sociedad civil y las comunidades. Partimos de la idea de que la evidencia adquiere valor público cuando no sólo es accesible, sino también comprensible, cuestionable y accionable por la ciudadanía, y planteamos que es necesario pasar de un enfoque de políticas basadas en evidencia (*evidence-based policy*, EBP) a uno de buena gobernanza de la evidencia; es decir, a la implementación de un conjunto de reglas, capacidades y salvaguardas que garanticen la apertura, la participación de la comunidad y la integridad de las políticas. Nuestro argumento plantea que no basta con obtener evidencia e información de calidad, y que es necesaria una definición adecuada sobre qué evidencia cuenta, para quién y con qué consecuencias.

Para desarrollar este argumento, hemos organizado el artículo de la manera siguiente. Una vez definido el concepto de bienes y valor públicos, y las razones que justifican la propuesta de la evidencia como un bien público, describimos brevemente la relación

y las diferencias entre transparencia, rendición de cuentas y evaluación. Posteriormente, enlistamos las condiciones que permiten que la evaluación cumpla su propósito público y argumentamos a favor de pasar de un presupuesto basado en resultados (PBR) a la gobernanza de la evidencia. A partir de lo anterior, elaboramos una tipología para clasificar los sistemas de evaluación y sus limitaciones. Finalmente, concluimos subrayando la necesidad de una gobernanza sólida y la mediación pública de resultados para evitar lecturas triunfalistas o instrumentalizaciones políticas de la evidencia. Los fines últimos de este trabajo son, en primer lugar, proponer un marco analítico que permita identificar de manera sistemática los casos en que las evaluaciones de programas sociales realmente funcionan como mecanismo de rendición de cuentas y mejora de la política social, y en segundo, argumentar a favor de avanzar del enfoque de la política basada en evidencia al de gobernanza de ella.

La evidencia como bien público

En economía, el bien público es un bien que se caracteriza por ser no excluible —no se puede impedir su uso a nadie— y no rival —su uso por una persona no reduce la posibilidad de que otra lo use—. La evidencia y el conocimiento se pueden compartir sin agotarse, por lo que son bienes no rivales. Sin embargo, puesto que son generados por individuos o instituciones específicas, se puede restringir su acceso. Por eso estos bienes no pueden mantener su condición como bien público por sí mismos; en los hechos suelen operar como bienes privados o, más precisamente, como bienes tipo club (excluíbles, aunque no necesariamente rivales).

La información surge de la observación y el procesamiento individual, y su difusión depende de quien la produce. Por sí sola, no es pública por naturaleza; incluso cuando se vuelve accesible, ello no garantiza que sea comprensible para otros. Además, el uso de un lenguaje técnico o de formatos inaccesibles puede convertirla en un bien reservado incluso bajo esquemas formales de transparencia. Para que la evidencia funcione como bien público se requieren dos condiciones: no exclusión efectiva y usabilidad social. Esto implica no sólo disponibilidad abierta, sino también una presentación comprensible y orientada a distintos usuarios.

En suma, debido a que la información no es intrínsecamente no excluible ni usable, se requiere intervención institucional para dotarla de estas dos cualidades y convertirla en

un bien público en sentido funcional. Pero ¿por qué queremos que la información y la evidencia tengan el carácter de un bien público? La literatura ofrece tres razones principales a favor de la evidencia como bien público, que sea democrática, instrumental y distributiva.

Razón democrática o el derecho a saber

Cuando la evidencia es financiada con recursos públicos, existe la obligación de dar razones públicas sobre su uso y sus resultados. La recomendación sobre ciencia abierta de la Organización de la Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO, 2021) propone un marco internacional para que el conocimiento financiado con fondos públicos sea abierto, accesible y reutilizable, bajo supervisión pública y con infraestructuras abiertas. En América Latina, gracias a la Corte Interamericana de Derechos Humanos (CIDH), se consolidó el acceso a la información pública como parte de la libertad de expresión (CIDH, 2006; United Nations, 1989). Este fundamento respalda que la evidencia de programas sociales sea accesible a la ciudadanía.

En México, el uso de la evidencia como bien público no es sólo deseable sino exigible. El artículo 6º constitucional garantiza el derecho al acceso a la información y establece el principio de máxima publicidad, el cual ordena divulgar información de interés público y justifica que la ciudadanía pueda escrutar, reutilizar y cuestionar la evidencia que respalda las decisiones de política social. Esto se explicita en la Ley General de Transparencia y Acceso a la Información Pública (LGTAIPI), la cual ordena que la información produzca «conocimiento público útil» y se publique en formatos accesibles, comprensibles y verificables (artículos 52 y 57), y hace obligatoria la divulgación de evaluaciones de programas, encuestas y estudios realizados con recursos públicos (artículo 65, fracciones XXXVIII-XXXIX), de manera responsable, con salvaguardas de privacidad y no revictimización.

Razón instrumental o la mejora en la toma de decisiones

Las personas pueden tomar mejores decisiones cuando tienen información adecuada; en tal sentido, ofrecer información clara y amplia mejora la calidad de vida de las personas. De ahí la importancia de presentar información diseñada para el usuario en general, es decir, «dirigida» a un amplio grupo de lectores, sin formalismos ni tecnicismos que excluyan a grupos específicos, y con una divulgación amplia. Esto hace de la información un instrumento para la toma de decisiones que incentiva cambios y mejoras basadas en evidencia.

Se ha demostrado que un diseño correcto de la estrategia para divulgar la información puede cambiar comportamientos sociales (OECD, 2017); pero si la información no es relevante o inteligible, los efectos se diluyen (Fox, 2015; Bovens, 2007). Por ello, la evidencia debe ir de la mano con la claridad o accesibilidad cognitiva (inteligibilidad) y con un diseño orientado al usuario. Esto implica tomar en cuenta las barreras sociotécnicas (estándares, metadatos, formatos y capacidades, entre otras cosas) que impiden el uso y la reutilización de la información (Janssen *et al.*, 2012; Zuiderwijk y Janssen, 2014).

De esta manera, podemos argumentar que cuando la evidencia se concibe como un bien público y facilita la inteligibilidad permite que se genere aprendizaje colectivo y mejoren las decisiones de política pública. Como se ha demostrado en diversos países, el acceso a la evidencia y la información favorece la innovación y la construcción de conocimiento, que a la vez mejora la eficiencia de las políticas públicas (OECD, 2015); el intercambio de saberes genera externalidades positivas, que justifican políticas de apertura (Stiglitz, 1999; UNESCO, 2021).

En suma, la evidencia clara e inteligible facilita la observación de la gestión pública, el control social y el aprendizaje institucional, que son incentivos para una mejora gubernamental.

Razón distributiva o la «política» de la evidencia

Decidir qué se mide, cómo se operacionaliza y para qué se usa la evidencia define la realidad pública de las políticas; fija problemas, parámetros de éxito y umbrales de intervención, y con ello redistribuye poder entre el Estado, las contralorías sociales y las comunidades. Como advierten la política de la evidencia y la coproducción entre conocimiento y orden político, los hechos públicos no «hablan solos», están configurados por reglas, capacidades e incentivos de gobernanza del saber; por eso la medición no es neutra, sino una tecnología del gobierno que puede legitimar o cuestionar decisiones (Parkhurst, 2017; Jasanoff, 2004; Porter, 1995). El caso mexicano muestra cómo estas decisiones de medición reconfiguran la «realidad» pública del problema; por ejemplo, controversias sobre si un programa combate el «hambre» o las «carencia por acceso a la alimentación» obligan a traducir lenguajes políticos a indicadores oficiales, lo que desplaza (o reencuadra) la discusión pública hacia categorías técnicas y umbrales específicos (Graizbord, 2024).

Esta redistribución es resultado de al menos tres mecanismos. El primero consiste

en visibilizar y construir la agenda pública; como argumentó Porter en un primer momento, y posteriormente Jasanoff, lo que se mide adquiere «existencia pública» y tiene acceso a recursos; lo que no se mide no se observa, queda fuera del radar, y por lo tanto es invisibilizado (Jasanoff, 2004; Porter, 1995). El segundo mecanismo tiene que ver con la posibilidad de hacer comparables y medibles cosas distintas. Al traducir realidades heterogéneas a métricas comparables se puede reducir la complejidad de los asuntos públicos y reordenar prioridades. Esta existencia de métricas también incide en el comportamiento de los actores que quieren alcanzar correctamente la meta impuesta (Espeland y Sauder, 2007).

Lo anterior también tiene que ver con el tercer mecanismo que opera en la redistribución del poder asociado a la evidencia: la gobernanza por indicadores. Ésta aparece cuando un número diseñado para medir termina dirigiendo la acción pública; es decir, de ser una medida pasa a ser un objetivo, lo que genera comportamientos estratégicos o distorsiones, pues se optimiza el indicador en lugar de optimizar el bienestar. El resultado suele tener efectos distributivos —cuando reciben recursos, prestigio o sanciones— y reputacionales —quién se muestra competente— en el sistema público (Campbell, 1976; Power, 1997).

En la práctica, quien diseña la métrica y fija el umbral decide qué cuenta como éxito y quién es prioritario y quién queda fuera de la prioridad. Una modificación en el indicador o sus ponderaciones puede mover flujos de recursos entre territorios o dependencias y redefinir qué aparece como «buen desempeño». Por ello, la política de la evidencia exige justificar públicamente qué, cómo y para qué se mide, lo que abre datos, métodos y toma de decisiones al escrutinio y la coproducción (Jasanoff, 2004; Porter, 1995); de lo contrario, la medición funcionará como tecnología de gobierno y no como actividad neutra.

Límites y excepciones a la apertura total

Concebir la evidencia como bien público no equivale a una apertura irrestricta. En México, la apertura está regida por el principio de máxima publicidad (Constitución Política de los Estados Unidos Mexicanos [CPEUM], artículo 6º) y, a la vez, por las pruebas de daño y de interés público, previstas en la LGTAIP, que obligan a ponderar beneficios y riesgos antes de difundir información (CPEUM, 2024; LGTAIP, 2025). En términos operativos, ello supone someter la publicación de evidencia a una prueba de interés público versus

riesgo de daño, documentar razones públicas cuando existan exclusiones justificadas y temporales, y fijar plazos de reserva y desclasificación, con la intervención del comité de transparencia y la elaboración de versiones públicas cuando proceda (LGTAIP, 2025; INAI, 2021).

Por definición, un bien público no se agota con el uso. Pero información divulgada sin salvaguardas puede perjudicar a personas o colectivos, causando discriminación, estigmatización territorial o reidentificación, por mencionar algunos efectos negativos. Por eso, en México el marco legal articula una apertura responsable con protección de datos personales y confidencialidad estadística. La Ley General de Protección de Datos Personales en Posesión de Sujetos Obligados (LGPDPPO, 2025) establece principios de licitud, finalidad, proporcionalidad, calidad y responsabilidad, medidas de seguridad y reglas reforzadas para datos sensibles. A su vez, la Ley del Sistema Nacional de Información Estadística y Geográfica (LSNIEG, 2021, 2025) declara que son estrictamente confidenciales los datos de informantes y habilita el acceso a microdatos anonimizados mediante controles y procedimientos de divulgación estadística. En la práctica esto se traduce en escalas de acceso y técnicas de anonimización, agregación y revisión de salidas, que se pueden complementar con buenas prácticas internacionales como el enfoque de «cinco seguros» o técnicas formales (por ejemplo, privacidad diferencial) cuando el riesgo es alto (Desai *et al.*, 2016; Dwork y Roth, 2017). En un contexto de evaluación de programas sociales esto significa evitar desagregaciones que identifiquen comunidades pequeñas o situaciones sensibles, publicar versiones agregadas cuando proceda o establecer canales de acceso controlado para investigación legítima, con trazabilidad y compromisos de uso responsable (LGTAIP, 2025; LGPDPSO, 2025; INAI, 2021).

Por lo tanto, la evidencia financiada con recursos públicos debe publicarse de forma inteligible y reutilizable. Pero su divulgación debe superar pruebas de necesidad y proporcionalidad, incluir gestión de riesgos de daño y apoyarse en mecanismos técnicos y organizacionales que protejan a personas y colectivos, en especial a niños, niñas y adolescentes. Así, la evidencia conserva su carácter de bien público (no excluyente y socialmente utilizable) sin sacrificar derechos ni la seguridad.

Transparencia, rendición de cuentas y evaluación: relaciones y diferencias

Antes de analizar los pasos que se deben dar para lograr el propósito público de la evi-

dencia es importante separar con claridad tres funciones: transparencia (disponibilidad y accesibilidad de información de interés público), rendición de cuentas (explicar y asumir la responsabilidad por decisiones y resultados ante instancias con capacidad de control y eventual sanción) y evaluación (mecanismo de aprendizaje y mejora de política, no un fin punitivo en sí mismo). Esta distinción continúa presente en la discusión académica nacional entre autores como Mauricio Merino, Sergio López Ayllón, Guillermo Cejudo, María del Carmen Pardo y José Antonio Caballero, entre otros; pero existe consenso en que la rendición de cuentas es un mecanismo de control que necesita articular normas, actores y procedimientos —información, justificación y consecuencias—, mientras que la transparencia funciona como un engrane del mecanismo que sin el resto de su arquitectura no garantiza la rendición de cuentas (Merino *et al.*, 2010).

Luis F. Aguilar ha mencionado que la evaluación forma parte del ciclo de política pública como un insumo para aprender y corregir, en tensión creativa con lo político; su valor depende de que la evidencia sea inteligible y se use para mejorar la toma de decisiones, no sólo para «mostrar cumplimiento» (Aguilar, 2017). En la misma línea, Cabrero (2005) subraya que la acción pública y la gestión deben entenderse en clave de coproducción con la sociedad. En este contexto, evaluar es una herramienta para mejorar la calidad de esa acción, no un trámite para palomear lo que sí hemos cumplido. La información es insumo y resultado de la evaluación; el acceso a ella, la transparencia, es un elemento necesario que permite la evaluación y mejora la gestión y la acción pública.

En México, la matriz de indicadores para resultados (MIR) y el presupuesto basado en resultados (PBR) han profesionalizado la planeación, pero también han facilitado rituales de cumplimiento (verificaciones o palomeos) que incentivan conductas defensivas en lugar de aprendizaje organizacional. Desde hace dos décadas, Arellano y Ramírez (2000) advirtieron los dilemas del gerencialismo al afirmar que cuando la métrica se vuelve meta se corre el riesgo de reducir la evaluación a un control formalista y de ignorar la incertidumbre y ambigüedad inherentes a los resultados públicos. Estudios recientes también documentan limitantes normativas, técnicas e ideológicas para consolidar el PBR, que favorecen el «palomeo» sobre el uso sustantivo de los hallazgos (Silva, 2024).

Graizbord (2024) documenta que, en la práctica, buena parte del esfuerzo evaluativo se concentra en «estandarizar» programas y narrativas para hacerlos medibles y comparables. Esto puede fortalecer la rendición de cuentas, pero también producir ejercicios

de «evaluación en papel» si los incentivos institucionales privilegian la demostración de cumplimiento sobre el aprendizaje y la corrección adaptativa. Esto refuerza la necesidad de que la transparencia se acompañe de rutas de uso y de consecuencias verificables.

La propia metodología propuesta a nivel nacional reconoce que no pueden aislarse la evaluación de los resultados y el impacto del diseño y la implementación. Las evaluaciones de consistencia y resultados (ECYR) del extinto Consejo Nacional de Evaluación de la Política de Desarrollo Social (Coneval) exigían revisar la lógica de intervención, la vinculación con la planeación y operación antes de juzgar el desempeño; y las guías de Secretaría de Hacienda y Crédito Público (SHCP) para la MIR insisten en una teoría de cambio explícita y trazabilidad de indicadores (Coneval, s.f.; SHCP, 2016). En términos de gobernanza, como plantea Cejudo (2013), la planeación es el puente entre evaluación y rendición de cuentas: sin claridad de fines e instrumentos, la evaluación se vuelve mero reporte.

Por lo tanto, la evaluación sólo coadyuva a la transparencia y rendición de cuentas cuando se plantea en un marco de reglas que aseguran apertura usable, independencia y uso efectivo de los hallazgos. Pero su fin propio es producir aprendizaje público y mejorar las políticas. En esta lógica, la gobernanza de la evidencia desplaza rituales de cumplimiento y reorienta instrumentos como la MIR y el PBR/SED hacia la explicación causal, la corrección adaptativa y la legitimidad democrática de las decisiones, articulando así los componentes de la rendición de cuentas —información, justificación y consecuencias— que la literatura ha destacado (Merino *et al.*, 2010; Aguilar, 2017; Arellano y Ramírez, 2000; Cabrero, 2005). Así, gobernada con las salvaguardas correspondientes, la evaluación deja de ser un trámite reservado y se aproxima al ideal de bien público de ser accesible, comprensible y útil para la ciudadanía.

Condiciones para que la evaluación cumpla su propósito público

Para garantizar que la evaluación sea un bien público, proponemos cuatro condiciones básicas: diseño desde el origen del programa, evaluación integral, rutas institucionales de uso de la evidencia y desarrollo de capacidades. Estas cuatro condiciones se derivan de dos fuentes: 1) la literatura sobre la evaluabilidad y el uso de la evaluación, la cual subraya que sin diseño, datos ni capacidad institucional la evidencia no se puede producir ni aprovechar, y 2) la perspectiva de gobernanza de la evidencia, que enfatiza reglas y

consecuencias para que el conocimiento no sea meramente simbólico. Son condiciones mínimas necesarias, pero no suficientes, para que la evaluación tenga valor público, es decir, que sea producida con trazabilidad, interpretada con criterios públicos y utilizada para decidir.

Diseño desde el origen del programa

Incorporar la evaluación desde que nace el programa implica construir y hacer explícita la teoría de cambio y la matriz de diseño que vincula necesidades, insumos, procesos y resultados esperados antes de implementar el programa; establecer preguntas de evaluación y criterios de éxito *ex ante*, y asegurar que los sistemas de información y las líneas de base permitan medir aquello en lo que se busca incidir (Rossi *et al.*, 2004; Funnell y Rogers, 2011). Además, es importante que antes de hacer una evaluación de impacto se mida la evaluabilidad del programa. Esto ayuda a saber si está en condiciones de ser evaluado para no hacer evaluaciones costosas que no pueden responder preguntas relevantes (Wholey, 2010). Incluir la evaluación desde el diseño también ayuda a especificar previamente indicadores, métodos y umbrales de decisión, lo que reduce discrecionalidades y el riesgo de sesgos *ex post*.

Desde la perspectiva de la «evaluabilidad» como proceso institucional, no basta contar con una teoría de cambio en abstracto, se requiere estabilizar definiciones operativas (población objetivo, criterios de elegibilidad, componentes, supuestos), producir documentos base (por ejemplo, notas conceptuales, reglas y manuales operativos) y acordar indicadores coherentes con tales conceptos. De lo contrario, los evaluadores terminan «reconstruyendo» *ex post* la lógica del programa para medirlo, lo que abre espacio a disputas políticas sobre qué cuenta como problema y qué como éxito (Graizbord, 2024). Además, el diseño debe incluir una arquitectura mínima de datos: plan de gestión de datos, responsables, diccionarios de variables, metadatos y procedimientos de calidad. Graizbord (2024) considera que sin estos insumos la información suele llegar a los equipos evaluadores como bases «crudas» desorganizadas o, por el contrario, como datos ya procesados sin trazabilidad, lo que impide replicar resultados y vuelve inestable la evidencia pública.

Evaluación integral

Evaluar sólo resultados o impactos no es suficiente cuando los programas son comple-

jos y sensibles al contexto. Se requieren evaluaciones integrales, que abarquen diseño, implementación (procesos y fidelidad), operación y resultados, y que tomen en cuenta distintos modelos de análisis, ya que de esta manera podemos ver la evidencia desde distintas perspectivas. Por ejemplo, modelos como contexto, insumos, procesos y productos (CIPP) orientan a examinar éstos de manera sistemática (Stufflebeam y Coryn, 2014); la evaluación realista propone entender qué funciona, para quién, en qué contextos y por qué mediante relaciones de contexto-mecanismo-resultado (Pawson y Tilley, 1997); y la ciencia de la implementación ofrece marcos para analizar determinantes y resultados de implementación (por ejemplo, CFIR) y *outcomes* específicos como aceptabilidad, adopción o sostenibilidad. Todos estos modelos buscan comprender los procesos que permiten interpretar efectos, identificar cuellos de botella y evitar diagnósticos erróneos del tipo «no funcionó, quítese», cuando en realidad falló la ruta de implementación. En suma, una evaluación pública robusta no separa efectos de mecanismos y contextos.

En este punto es útil distinguir entre 1) evaluaciones que «hacen evaluable» el programa (diagnósticos de diseño, clarificación de indicadores, revisión de operación y coordinación interinstitucional), y 2) evaluaciones que estiman efectos. La presión por demostrar «impacto» puede generar una suerte de «impactomanía» —la idea de que sólo los métodos experimentales validan una política—; pero la evidencia etnográfica sugiere que en burocracias reales el cuello de botella suele estar en la operación, la disponibilidad de datos y la alineación de incentivos, por lo que las evaluaciones de proceso y de diseño con frecuencia son condiciones de posibilidad para cualquier evaluación de resultados (Graizbord, 2024).

Rutas institucionales de uso de la evidencia

La evaluación sólo genera valor público cuando se usa para decidir. La literatura distingue usos instrumentales (cambios concretos de diseño o presupuesto), conceptuales (aprendizaje y cambio de marcos mentales) y simbólicos (legitimación) (Henry y Mark, 2003). Para potenciar los dos primeros y evitar que el tercero sea el dominante, la evaluación debe diseñarse con un enfoque de uso: identificar desde el inicio quién decide qué, con qué evidencia mínima y en qué ventana temporal; producir paquetes de datos e información que puedan reproducirse y comprenderse, y acompañar los hallazgos con planes de adopción que vinculen hallazgo-acción-responsable-plazo y esquemas «cumplir o expli-

car» cuando no se adopten recomendaciones (Patton, 2008; Nutley *et al.*, 2007). Además, los sistemas públicos necesitan bucles de aprendizaje (revisiones a tres, seis y 12 meses, tableros de seguimiento, pruebas piloto) que conviertan la evidencia en mejora continua y eviten la «transparencia ritual», en la que se publican informes sin consecuencias (Bovens, 2007).

Desarrollo de capacidades

Para que lo anterior sea sostenible se requieren capacidades organizacionales y reglas claras. La evaluación como función debe contar con personal formado en métodos (mixtos), gestión de datos, ética y comunicación pública; protocolos y estándares (por ejemplo, criterios OECD/DAC de 2019, principios FAIR de datos, guías TOP de apertura reproducible), y rutas institucionales de uso que articulen la evaluación con planeación, presupuesto y regulación (Nosek, 2015; Wilkinson *et al.*, 2016; OECD, 2019). La construcción de capacidad evaluativa —formación, acompañamiento, incentivos y gobernanza interna— es una condición para que la evidencia se genere con calidad y se utilice; la literatura muestra que las organizaciones que invierten deliberadamente en estas capacidades aumentan la probabilidad de uso y la calidad de las decisiones (Labin *et al.*, 2012; Preskill y Boyle, 2008). Finalmente, cuando es difícil la atribución directa, enfoques como el análisis de contribución ayudan a establecer cadenas causales plausibles útiles para decidir en entornos complejos, lo que evita el dogmatismo y el escepticismo paralizante.

En conjunto, estas condiciones permiten alinear la función evaluativa con su propósito público: producir evidencia creíble, inteligible y utilizable para la rendición de cuentas y la mejora de la acción gubernamental. Sin ellas la evaluación corre el riesgo de convertirse en costo hundido y ritual de cumplimiento, con ellas se integra a la gobernanza de la evidencia y contribuye a su tratamiento como bien público. Como argumenta Graizbord (2024), el «propósito público» de la evaluación depende de arreglos sociotécnicos que permiten: 1) producir datos y documentos confiables; 2) coordinar relaciones entre operadores, técnicos y evaluadores externos, y 3) traducir hallazgos a un lenguaje público sin perder rigor, lo que evita el secretismo y la «transparencia ritual».

De las políticas basadas en evidencia a la gobernanza de ella

Como hemos argumentado, la información y los resultados de una evaluación no son

intrínsecamente bienes públicos, se requieren reglas que garanticen la no exclusión y un uso social efectivo. Por lo tanto, no basta la evidencia como base de la política; con esta lógica es necesario transitar a la gobernanza de la evidencia, entendida como el conjunto de reglas, capacidades y la independencia que aseguran que lo que cuenta como «evidencia» sea representativo, verificable y utilizable, lo que habilita la rendición de cuentas y el aprendizaje público (Parkhurst, 2017; Fung *et al.*, 2007; Fox, 2015). Desde este enfoque, la evaluación pasa de ser un proceso de «palomeo» a ser un mecanismo de aprendizaje, en el cual no caben veredictos binarios sobre lo que funciona y lo que no. Su función consiste en explicar por qué y cómo se producen los resultados observados, independientemente de que se utilicen métodos cuantitativos, cualitativos o mixtos. Sólo entendiendo los procesos que determinan la relación de causalidad se pueden mejorar las decisiones y acciones públicas (Bovens, 2007).

Para aplicar el criterio de gobernanza de la evidencia se deben utilizar criterios públicos para ordenar y definir qué se mide, cómo se produce y transparenta, quién participa en su definición e interpretación y cómo se usa para decidir. Nosotros proponemos los cuatro elementos básicos que se describen en los cuatro subapartados siguientes.

Apertura proactiva de información reproducible

No basta publicar informes inteligibles, una buena gobernanza requiere la publicación por defecto de toda la información utilizada para que los hallazgos puedan ser replicados: datos (protegidos garantizando la confidencialidad), diccionarios de variables, metadatos, códigos, protocolos de investigación, entrevistas, etcétera. Además de resúmenes ejecutivos, gráficas y todos aquellos elementos que permitan visualizar y leer fácilmente la información. Esto permite cumplir con los principales lineamientos de transparencia establecidos a nivel internacional y los principios FAIR, pues son fáciles de encontrar, accesibles para la mayoría de las personas, interoperables —es decir, que datos y resultados se puedan trabajar fácilmente con otros sistemas, herramientas y conjuntos de datos, lo cual implica que las bases de datos no son publicaciones PDF escaneadas o archivos con candados— y reutilizables (Wilkinson *et al.*, 2016; Nosek, 2015). De esta manera la evidencia es verificable y susceptible de ser reutilizada socialmente. En todos los casos se requiere una apertura responsable, a partir de condiciones de acceso a los datos que permitan maximizar el valor y reducir al mínimo el riesgo de su publicación (Desai *et al.*,

2016); no se pueden utilizar los mismos mecanismos de divulgación para todos los datos; en el caso de las infancias, como ya se mencionó, se requieren salvaguardas particulares, ya que los riesgos son mayores.

Participación sustantiva de actores afectados

Muchas veces se confunde socialización *ex post* con participación efectiva. La participación implica un poder compartido entre los distintos actores afectados para definir preguntas, indicadores, supuestos e interpretación a lo largo del ciclo de producción de la evidencia. Incluir a comunidades y organizaciones mejora la pertinencia, legitimidad y probabilidad de uso. Esto implica apertura del proceso a las personas afectadas, la negociación para definir objetivos, métodos, interpretación, intercambio y difusión de resultados, y la flexibilidad para adaptarse de manera constante al entorno y las necesidades que van surgiendo durante el propio proceso de investigación (UNESCO, 2010; Murrieta *et al.*, 2014).

Para que la generación de evidencia sea pertinente y de calidad es importante incluir en el proceso una diversidad de participantes, garantizar la presencia de grupos minoritarios o con menor poder y facilitar las condiciones para que las relaciones de poder entre los participantes no inhiban la participación de algunos. El diálogo y la deliberación entre los grupos también requiere que se tomen en cuenta la motivación, el acceso a la información y la formación de los participantes. En muchos casos esto implica formar a las personas involucradas antes de iniciar el proceso participativo (Planas *et al.*, 2014). En todos los casos el evaluador debe desempeñar un papel activo como facilitador y conocer el contexto en el que se realiza el proceso de evaluación o investigación.

Integridad e independencia en el proceso de generación de información

Asegurar la integridad e independencia en la generación de evidencia implica blindar el proceso ante posibles sesgos operativos, presiones políticas e incentivos editoriales. La integridad se ocupa de cómo se produce el conocimiento —esto es, de la transparencia metodológica y analítica—, mientras que la independencia se refiere a quién lo produce y quién es el árbitro en las controversias. En la práctica, esto exige reglas claras de conflicto de interés (declaración *ex ante* de financiamiento, relaciones, posibles variantes y su gestión efectiva), así como el registro de protocolos y planes de análisis especificados previa-

mente que señalen cuándo se trata de análisis exploratorio o de confirmación de hipótesis; que enlisten hipótesis, resultados esperados, subgrupos, poder estadístico y modelos previstos antes de analizar los datos. Cuando el curso del estudio exige cambios, un registro de desviaciones transparente preserva la trazabilidad y permite evaluar la robustez de los hallazgos (Nosek *et al.*, 2015, 2018).

Para contrarrestar el sesgo de publicación y el «*file-drawer problem*» (problema del cajón), que conduce a la subrepresentación de resultados nulos o negativos (Rosenthal, 1979; Franco *et al.*, 2014), la gobernanza debe incorporar la obligación de divulgar este tipo de resultados, cuando sea posible mediante el modelo de *registered reports* o artículos registrados, en que las revistas científicas evalúan y aceptan los estudios antes de conocer sus resultados con base en la solidez metodológica, lo que garantiza la publicación de evidencia válida aunque los hallazgos sean nulos o negativos (Chambers, 2013). Cuando surgen controversias sobre diseño, análisis o interpretación, el arbitraje metodológico —paneles externos, revisión ciega, auditorías de código y datos, y dictámenes públicos— agrega garantías de imparcialidad y reproducibilidad. Finalmente, la separación funcional entre quienes operan programas y aquellos que evalúan, reforzada con comités metodológicos dotados de mandato público, criterios de idoneidad e incompatibilidades, reduce los incentivos a la captura y fortalece la credibilidad del sistema (Parkhurst, 2017). Con este andamiaje, la evidencia deja de depender de la autoridad de quien la produce y se legitima en la verificabilidad del proceso.

Salvaguarda de derechos

La apertura responsable requiere acciones que maximicen el valor público sin producir daño. En todos los casos es fundamental respetar los principios de privacidad y no revictimización, respetando el uso de consentimientos informados y manejando de manera adecuada los microdatos. Antes de publicar microdatos o desagregaciones finas es imprescindible hacer evaluaciones de impacto en derechos que ponderen riesgos de estigmatización territorial, reidentificación y exposición a violencia, y que definan medidas de mitigación y establezcan responsables. Cuando conviven el interés público y riesgos elevados —por ejemplo, en investigaciones en curso—, deben aplicarse exclusiones justificadas y temporales con razones públicas y horizontes de desclasificación explícitos ofreciendo, mientras tanto, versiones agregadas para preservar la utilidad social.

Estas salvaguardas descansan en una combinación de técnicas y arreglos organizacionales. Entre las primeras destacan el anonimato o uso de seudónimos, la agregación geográfica o temporal, la supresión de variables con información escasa, la codificación del participante y, cuando el riesgo lo justifica, mecanismos formales como la privacidad diferencial; entre los segundos, el uso de entornos seguros de acceso y la revisión de salidas antes de su liberación pública, conforme al marco de los «cinco seguros» (proyecto, personas, datos, entornos y salidas), que equilibra utilidad y protección (Desai *et al.*, 2016). Además, un diseño comunicativo inclusivo —lenguaje y visualizaciones que eviten estigmas, materiales de lectura fácil, formatos culturalmente pertinentes y apoyo psicosocial cuando el trabajo de campo lo requiera— reduce el riesgo de daño simbólico y favorece el uso social de la evidencia (UNICEF, 2025).

Por lo tanto, sin apertura proactiva e inteligibilidad, participación sustantiva, integridad e independencia y salvaguardas de derechos, la evidencia permanece excluíble —reservada a quien paga, autoriza o controla— incluso en contextos con leyes de transparencia. Con estas condiciones, en cambio, la evidencia se aproxima a un estatus funcional de bien público: accesible, verificable, contestable y realmente usable para la rendición de cuentas y la mejora de las políticas.

Construcción de la evidencia como bien público

Hemos argumentado que la evidencia adquiere las propiedades de un bien público sólo cuando cumple simultáneamente con cuatro condiciones: no exclusión (apertura usable) y usabilidad social (inteligibilidad), gobernanza efectiva (reglas, capacidades e independencia), participación de comunidades afectadas y salvaguardas de derechos. La construcción de esta evidencia requiere una arquitectura sustentada en cuatro elementos conceptuales.

El primer elemento, los cimientos sobre los cuales se debe construir esta estructura, consiste en distinguir con claridad entre transparencia, rendición de cuentas y evaluación. La transparencia aporta información, la rendición de cuentas exige justificación y consecuencias, y la evaluación es un mecanismo de aprendizaje y mejora continua (Merino *et al.*, 2010; Aguilar, 2017). Sin esta distinción conceptual no existen cimientos sólidos para un sistema evaluativo que genere valor público. El segundo elemento exige transitar de políticas basadas en evidencia a una gobernanza de la evidencia, entendida como el conjunto de reglas y capacidades que aseguran la integridad, la independencia y un uso

efectivo de la información generada. El tercer elemento reconoce que la apertura genera valor sólo si es usable y hace referencia a la «transparencia dirigida» y los principios FAIR; esto implica que la información debe ser accesible y publicarse en formatos comprensibles y reutilizables, de lo contrario la apertura se convierte en un gesto formal sin impacto social (Fung *et al.*, 2007; Wilkinson *et al.*, 2016; Zuiderwijk y Janssen, 2014). Finalmente, el cuarto elemento busca evitar el ritualismo de los indicadores, es decir, la tendencia a convertir los sistemas de monitoreo en fines en sí mismos, sin aprendizaje ni consecuencias reales, lo cual hace que la medición degenera en «palomeo» burocrático y pierda legitimidad.

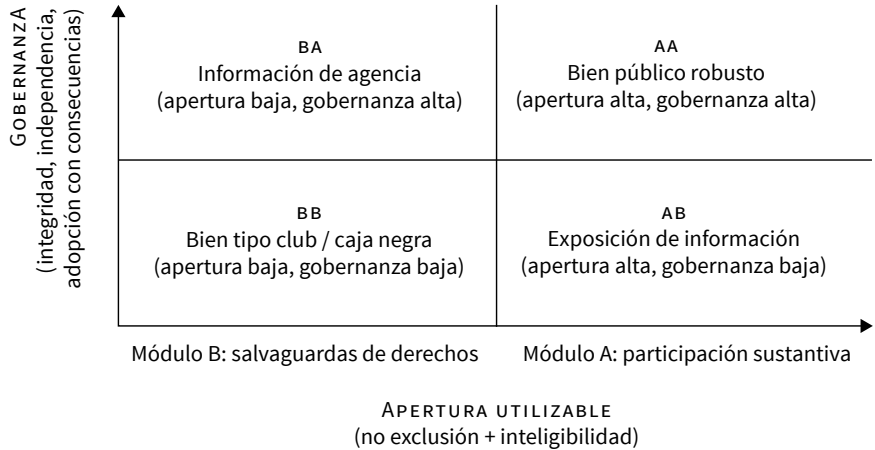
A partir de estos fundamentos surge la pregunta clave: ¿qué tan cerca está un sistema de producir y usar evidencia con esas propiedades públicas? Para responderla, proponemos una tipología analítica que permite clasificar los sistemas de evaluación no para sancionar, sino para orientar transformaciones que los acerquen al ideal de un bien público robusto: evidencia accesible, verificable, contestable y útil para la toma de decisiones, especialmente en el ámbito de las políticas para la infancia.

Matriz de gobernanza de la evidencia

Una vez establecidos los fundamentos conceptuales de la evidencia como bien público, la siguiente tarea consiste en operativizar tales condiciones en un esquema que permita distinguir distintos grados de desarrollo institucional. Para facilitar esta operativización desarrollamos una matriz de gobernanza de la evidencia (véase el esquema 1). Con ello buscamos ofrecer una herramienta analítica para ubicar los sistemas de evaluación según su capacidad de generar valor público.

El diseño de la matriz de gobernanza de la evidencia parte de dos ejes principales. El eje horizontal (x) representa la apertura utilizable, entendida como la combinación de no exclusión e inteligibilidad de la información. Este eje abarca un continuo que va desde la publicación escasa o ritual hasta la apertura proactiva y reproducible, con informes comprensibles, microdatos, metadatos y código en formatos abiertos, entre otras cosas. El eje vertical (y) mide la gobernanza para el uso, que combina integridad, independencia y adopción con consecuencias; incluye prácticas como el prerregistro, la gestión de conflictos de interés, la publicación de resultados negativos, el arbitraje metodológico y los mecanismos de cumplimiento o explicación.

Esquema 1 | Matriz de gobernanza de la evidencia



Fuente: Elaboración propia.

A partir de la interacción entre ambos ejes de la matriz del esquema 1, en este trabajo proponemos una tipología con cuatro categorías para clasificar sistemas de evidencia:

1. *Bien público robusto* (cuadrante AA, apertura y gobernanza altas). Apertura reproducible e independencia con uso verificable. La evaluación se orienta al aprendizaje público y refuerza la legitimidad democrática.
2. *Exposición de información* (cuadrante AB, apertura alta y gobernanza baja). Abundante publicación sin consecuencias. La transparencia se convierte en ritual y la evidencia pierde capacidad de transformación.
3. *Información de agencia* (cuadrante BA, apertura baja y gobernanza alta). Uso interno tecnocrático con poca apertura; es decir, información que genera u obtiene una institución y no comparte con el público en general u otras agencias (instituciones). Eficacia operativa, pero déficit de legitimidad social.
4. *Club o caja negra* (cuadrante BB, apertura y gobernanza bajas). Evidencia excluible, sin uso ni control. Es el extremo opuesto del bien público.

A lo largo de estos cuadrantes operan dos módulos transversales o condiciones habilitantes. El módulo A, de participación sustantiva, evalúa en qué medida comunidades y sociedad civil codefinen preguntas, indicadores e interpretación de resultados, en lugar de

ser sólo informadas *ex post*. El módulo B, de salvaguardas de derechos, mide la existencia y efectividad de principios de privacidad, no daño y no revictimización, particularmente en la producción y difusión de evidencia sobre niñas, niños y adolescentes (CPEUM, artículo 6º; LGTAIP; LGPDPSO; Ley General de los Derechos de Niñas, Niños y Adolescentes [LGDNNA]). Ningún sistema puede considerarse un bien público en sentido pleno sin un nivel mínimo de protección en este módulo.

En síntesis, esta tipología convierte los principios normativos de la evidencia como bien público en una herramienta práctica de diagnóstico. Permite reconocer posiciones intermedias, orientar trayectorias de mejora (por ejemplo, de «exposición de información» a «bien público robusto») y ofrecer un lenguaje común para discutir la madurez de los sistemas de evaluación en términos de apertura, gobernanza, participación y derechos.

Proposiciones teóricas e implicaciones observables

Con la tipología ya definida, formulamos cinco proposiciones teóricas derivables que especifican mecanismos entre las dimensiones del modelo: apertura usable (X), gobernanza para el uso (Y), participación (A) y salvaguardas de derechos (B), y una serie de resultados intermedios esperados: uso social, legitimidad, pertinencia, usabilidad y confianza. Estas proposiciones no buscan anticipar impactos sustantivos de las políticas, sino poner a prueba si los arreglos institucionales que habilitan la evidencia como bien público generan los efectos intermedios que sugiere la literatura (por ejemplo, transparencia dirigida, gobernanza de la evidencia, participación sustantiva y apertura responsable) (Fung *et al.*, 2007; Fox, 2015; Parkhurst, 2017). Cada proposición es observable y falsable; se vincula con indicadores de los ejes y módulos (X, Y, A, B) y con criterios externos de verificación (adopción de recomendaciones, tiempos hallazgo-acción, litigios, reanálisis de datos), con lo que se prepara el terreno para la operacionalización de la matriz. A continuación, enunciamos estas proposiciones, con algunas implicaciones observables.

P1 (exclusión → no uso). Cuando la evidencia se mantiene excluible (acceso restringido, ausencia de microdatos/código) disminuye la probabilidad de que la usen actores no gubernamentales, y con ello la capacidad de control democrático, aun en contextos con leyes de transparencia. Por lo tanto, podemos esperar una relación positiva entre inteligibilidad y usos sociales.

P2 (integridad → legitimidad). El prerregistro, la independencia y la publicación de

resultados (incluidos los nulos o negativos) aumentan la legitimidad de la evidencia porque hacen trazables las decisiones metodológicas, permiten la auditoría y reducen las sospechas de sesgo o captura. Cuando existen riesgos de sesgo de selección o se observan efectos heterogéneos —algo frecuente en políticas complejas— estas prácticas ayudan a reportar supuestos, límites e incertidumbre con claridad y a sostener controversias en términos públicos. Mayor integridad se asocia con menor conflictividad (menos aclaraciones/fe de erratas, menor discrepancia entre dictámenes externos), mayor aceptación por órganos de control y mayor probabilidad de que los hallazgos sean reutilizados en decisiones públicas.

P3 (participación → pertinencia). La participación sustantiva en el planteamiento de preguntas e indicadores mejora la pertinencia y aumenta la adopción de recomendaciones. Cuando la participación es alta (codiseño y validación con trazabilidad de cambios) aumenta la probabilidad de adopción con consecuencias y disminuye el tiempo entre generación de hallazgos y cambios.

P4 (traducción → usabilidad). La mediación pública (inteligibilidad) transforma la evidencia técnica en insumo accionable, especialmente en políticas de infancia, en que la lectura fácil y las visualizaciones no estigmatizantes amplían la base de usuarios. La inteligibilidad se asocia con un mayor número de decisiones formales vinculadas a la evidencia, más consultas ciudadanas sustantivas y una mayor cobertura mediática de calidad.

P5 (salvaguardas → confianza). Las salvaguardas de privacidad y no daño aumentan la confianza social, lo que evita retraimientos de datos, recursos legales o bloqueos a la apertura. Unos niveles de salvaguardas altos se asocian con menos litigios o recursos de revisión, mayor estabilidad del calendario de publicación y más solicitudes de acceso por parte de la academia y de organizaciones de la sociedad civil.

En conjunto, estas proposiciones delimitan una agenda de verificación empírica. Orientan la operacionalización en indicadores y escalas, y permiten contrastar, en distintos contextos y períodos, si los sistemas avanzan hacia el ideal de bien público robusto o permanecen en arreglos de exposición ritual o caja negra.

De la tipología a la operacionalización

Para que la matriz de gobernanza sea útil en la práctica es necesario traducir sus dimensiones en indicadores observables y escalas verificables. Buscamos generar una herramienta

que permita medir hasta qué punto los sistemas de evaluación reúnen los atributos de un bien público robusto. Con tal fin, convertimos los principios —apertura, inteligibilidad, integridad, independencia, participación y salvaguardas— en variables cuantificables que permiten comparar, monitorear avances y orientar intervenciones de mejora.

De esta forma, la matriz se transforma en un instrumento de diagnóstico dinámico que vincula la teoría con la práctica. Al asignar valores a cada dimensión y módulo se obtiene un mapa que no sólo posiciona los sistemas dentro de los cuadrantes ideales (bien público robusto, exposición de información, información de agencia, club/caja negra), sino que también identifica las palancas específicas de mejora para avanzar hacia una gobernanza de la evidencia más abierta, participativa y responsable.

Operacionalización: indicadores y escalas

La operacionalización propuesta se basa en un conjunto de indicadores específicos que permiten medir de manera sistemática y comparable el grado de apertura usable (eje x), la gobernanza para el uso (eje y) y los módulos transversales de participación sustantiva (A) y salvaguardas de derechos (B). Proponemos valorar cada indicador en una escala ordinal de 0 a 3, que refleja niveles de desarrollo institucional desde la ausencia de prácticas (0) hasta la adopción consolidada y verificada por terceros (3). Este enfoque combina el análisis normativo con la medición empírica, lo que permite situar los sistemas de evaluación dentro de los cuadrantes de la matriz y al mismo tiempo identificar brechas y trayectorias de mejora. Los indicadores que se presentan a continuación buscan capturar el equilibrio entre apertura y responsabilidad, y ofrecen una herramienta práctica para el monitoreo de la calidad democrática de los sistemas de evaluación, en especial de los relacionados con políticas de infancia y desarrollo social.

Eje x: apertura usable (no exclusión + inteligibilidad). El eje x mide el grado en que la información y la evidencia generadas por los sistemas de evaluación son públicamente accesibles, comprensibles y reutilizables. Este eje traduce la idea de no exclusión en criterios operativos: publicación proactiva y oportuna, disponibilidad de microdatos, metadatos y código reproducible, así como la inteligibilidad del contenido mediante resúmenes claros, visualizaciones y materiales adaptados a distintos públicos. Siguiendo la noción de transparencia dirigida (Fung *et al.*, 2007) y los principios FAIR (Wilkinson *et al.*, 2016), se busca que la información pueda ser entendida y reutilizada por actores diversos, incluyendo

academia, sociedad civil y comunidades locales. Por ello, el eje x no se limita a medir la cantidad de información publicada, también mide la calidad comunicativa y técnica de su apertura, y distingue entre transparencia ritual (publicación sin uso) y apertura significativa que habilita escrutinio, aprendizaje y participación social.

Con base en la literatura, proponemos al menos cuatro indicadores para medir la apertura usable: 1) *publicación proactiva* (porcentaje de evaluaciones publicadas dentro de un rango determinado de días tras la conclusión de la misma); 2) *reproducibilidad* (porcentaje de estudios con paquetes replicables —microdatos y códigos— verificados por terceros); 3) *inteligibilidad* (índice de lectura/inclusión de resúmenes en lectura fácil, glosarios y guías de interpretación), y 4) *adopción* (porcentaje de evaluaciones que derivan en rediseños o ajustes presupuestarios formalmente documentados y aplicación del criterio cumplimiento o explicación).

Eje y: gobernanza para el uso (integridad + independencia + adopción con consecuencias). Este eje evalúa la solidez institucional del proceso de generación y aplicación de evidencia. Mientras que la apertura usable se centra en el acceso, el eje y se enfoca en el modo en que la evidencia se produce, interpreta y utiliza en los sistemas públicos. La gobernanza efectiva exige integridad metodológica (transparencia en diseños, planes de análisis y manejo de conflictos de interés), independencia funcional (separación entre operación y evaluación, arbitraje técnico y revisión externa) y mecanismos que aseguren adopción con consecuencias, como decisiones formales derivadas de la evidencia o esquemas de cumplimiento o explicación. De esta forma, el eje y distingue entre los sistemas que generan evidencia como ejercicio burocrático y aquellos que la transforman en decisiones verificables y trazables, lo que cierra el ciclo de rendición de cuentas y aprendizaje público.

Una manera de medir la integridad es la existencia de prerregistros, metodologías rigurosas y la divulgación de resultados negativos, en un marco de reglas claras de respuesta ante cualquier resultado, lo que da lugar a cuatro variables: 1) *prerregistro* (porcentaje de evaluaciones con registro de protocolo y declaración de posibles conflictos de interés); 2) *rigor metodológico* (evaluaciones con objetivos, hipótesis y metodologías claras y bien articuladas); 3) *divulgación negativa* (proporción de informes con resultados no favorables, publicados en tiempo), y 4) *adopción con consecuencias* (porcentaje de decisiones formales tomadas a partir del cumplimiento o la justificación de los resultados).

Módulo A: participación sustantiva de actores afectados. El módulo de participación

sustantiva actúa como una dimensión transversal que examina el grado en que la evidencia incorpora voz y agencia de los actores afectados, más allá de la socialización *ex post* de los resultados. Inspirado en la literatura sobre evaluación participativa, este módulo analiza la presencia de mecanismos de codiseño de preguntas y criterios de éxito, validación pública de hallazgos y adopción de cambios sugeridos por los participantes.

La participación sustantiva no se reduce a consultas simbólicas, supone la negociación deliberativa del conocimiento y la incorporación de la diversidad de perspectivas. En términos normativos, este módulo introduce un principio de legitimidad epistémica: cuanto mayores son la participación y diversidad de actores, menor es el riesgo de captura tecnocrática y mayor el valor público de la evidencia. A partir de lo anterior, consideramos que deben incluirse al menos dos variables: 1) *codiseño* (porcentaje de evaluaciones con actas de consulta a grupos afectados y número de aportes incorporados), y 2) *validación pública* (número de audiencias/talleres de retroalimentación y evidencia de cambios derivados de dichas reuniones, o justificación cuando no los haya).

Módulo B: salvaguardas de derechos. El módulo de salvaguardas de derechos garantiza que la apertura de información y la generación de evidencia no produzcan daño a personas o colectivos. Retoma los principios de privacidad, consentimiento informado y no revictimización previstos en la CPEUM (artículos 4º y 6º) y en leyes mexicanas como la LGTAIP, la LGPDPPSO y la LGDNNA, así como en estándares internacionales de derechos de la niñez (United Nations, 1989; UNICEF, 2025). Este módulo evalúa la existencia y efectividad de mecanismos como pruebas de daño/interés público, evaluaciones de impacto en derechos, anonimización y revisión de salidas, y canales de acceso controlado bajo el enfoque de los cinco seguros (Desai *et al.*, 2016).

En este sentido, las salvaguardas no limitan la apertura, sino que la legitiman, al equilibrar la utilidad social de la evidencia con la protección de las personas. Sin niveles mínimos en este módulo, un sistema no puede considerarse un bien público robusto, por más avanzada que sea su apertura o su gobernanza técnica. Por lo tanto, son al menos tres los indicadores que deben utilizarse para medir la existencia de salvaguardas de derechos: 1) *privacidad* (existencia y cumplimiento de la evaluación de impacto en protección de datos); 2) *riesgo de privacidad* (tasa de reidentificación menor o igual a un umbral determinado en auditorías), y 3) *principio de no daño* (resoluciones de no publicación temporal con plazos y justificación, bien argumentadas).

Reglas de clasificación y ponderación

Una vez evaluados los indicadores de cada eje y cada módulo, es necesario establecer reglas de decisión coherentes que permitan sintetizar la información y clasificar los sistemas de evaluación en la matriz de gobernanza de la evidencia. La ponderación y los umbrales propuestos responden a un principio de equilibrio entre apertura y gobernanza: un sistema no puede considerarse un bien público robusto únicamente por cumplir en uno de los ejes. Así, los puntajes obtenidos en las dimensiones X (apertura usable) y Y (gobernanza para el uso) se combinan mediante promedios ponderados para determinar su posición en los cuadrantes, mientras que los módulos transversales A (participación sustantiva) y B (salvaguardas de derechos) actúan como condiciones de suficiencia.

Puesto que la clasificación no busca jerarquizar ni sancionar, las reglas de ponderación integran umbrales claros (por ejemplo, 1.5 para distinguir entre bajo y alto) y criterios de ajuste que previenen falsas clasificaciones: un sistema con alta apertura, pero sin protección de derechos, o con independencia, pero sin participación, no puede ser considerado plenamente público. En consecuencia, la combinación de ejes y módulos ofrece una lectura integral del grado en que un sistema materializa la evidencia como bien público, e identifica brechas y prioridades de acción.

Al asignar puntajes verificables a cada eje y cada módulo, los sistemas de evaluación pueden situarse en un mapa analítico que revela no sólo su nivel de apertura y gobernanza, sino también los déficits estructurales que limitan su valor público. Esta lectura comparativa ofrece evidencia accionable para decidir dónde invertir capacidades, ajustar reglas o fortalecer salvaguardas. Más que un instrumento de control, proponemos en este trabajo un instrumento que busca consolidar un aprendizaje institucional reflexivo, orientado a transformar los sistemas de evaluación en verdaderos bienes públicos: accesibles, verificables, contestables y útiles para la toma de decisiones.

Buena parte de las prácticas actuales miden y publican, pero no siempre aseguran inteligibilidad, participación significativa, independencia ni uso con consecuencias, y a menudo derivan en transparencia mecánica o «palomeos». La matriz propone un cambio de lógica: de tener indicadores cumplidos a gobernar la evidencia como bien público, integrando reproducibilidad, integridad, adopción con consecuencias y protecciones por diseño para niñas, niños y adolescentes (United Nations, 1989; Wilkinson *et al.*, 2016; Nosek *et al.*, 2015).

Consideramos que los resultados del uso de esta herramienta permitirán identificar qué tipo de sistema tenemos (exposición de información, información de agencia, club/caja negra o bien público robusto) y, sobre todo, qué palancas concretas pueden activarse para emigrar hacia un modelo de gobernanza de la evidencia que combine apertura responsable, independencia, participación y rendición de cuentas democrática. En el cuadro 1 (p. 34) ilustramos estos cuadrantes por medio de ejemplos hipotéticos y recomendaciones específicas para cada tipo de sistema.

Limitaciones y validez del modelo propuesto

Es importante subrayar que la matriz que proponemos en este trabajo tiene como fin ser un instrumento analítico que permita clasificar sistemas de evaluación o generación de evidencia. Su propósito es orientar la mejora institucional, no medir desempeño sustantivo. Como toda tipología, simplifica la realidad, por lo que sus resultados deben leerse como un mapa aproximado que ayuda a identificar fortalezas, brechas y posibles rutas de avance hacia una gobernanza más abierta y responsable de la evidencia. Esto plantea una serie de limitaciones, que enlistamos a continuación.

1. Conceptualizar la «evidencia como bien público» es complejo. Si no tomamos en cuenta esta complejidad y simplemente aplicamos el concepto sin anclar sus dimensiones corremos el riesgo de estirarlo a tal grado que pierda poder descriptivo. En este artículo sugerimos cuatro dimensiones: apertura usable, gobernanza para el uso, participación y salvaguardas. Al utilizarlas es importante hacer explícita la ponderación y tomar en cuenta que los indicadores pueden no ser equivalentes entre contextos institucionales. No es igual comparar sectores que estados o municipios, por lo que es fundamental documentar los supuestos y utilizar las herramientas estadísticas para validar la comparación. Esto se puede lograr partiendo de un marco conceptual sólido, de variables claramente definidas y haciendo explícitos los mecanismos operativos y los criterios de codificación.
2. Toda decisión sobre cómo se combinan los indicadores —por ejemplo, qué peso tiene cada dimensión o qué valor se considera «alto» o «bajo»— implica cierto grado de arbitrariedad. Pequeñas variaciones en los criterios pueden modificar la posición final de un sistema en la matriz. Por eso es importante reconocer que los resultados no son absolutos, sino aproximaciones útiles. Para fortalecer su confiabilidad se reco-

Cuadro 1. Ejemplos hipotéticos de uso de la matriz de gobernanza

Cuadrante / tipo	Descripción sintética	Puntajes típicos (X, Y, A, B)	Ejemplo ilustrativo hipotético (infancia)	Principales brechas
AA. Bien público robusto.	Apertura reproducible más uso con consecuencias; alta legitimidad.	X≈2.6 Y≈2.4 A≈1.8 B≈2.5	<i>Sistema estatal de evaluación de vacunación y primera infancia</i> : publica informes, microdatos y código; resoluciones de adopción con cumplimiento o explicación.	Consolidar resultados negativos; fortalecer metaevaluación y participación de cuidadores/niñas, niños y adolescentes.
AB. Exposición de información.	Mucha publicación, poco uso; riesgo de «palomeo».	X≈2.4 Y≈1.2 A≈1.0 B≈2.3	<i>Portal de becas escolares</i> : paneles de control y PDF, pero sin cambios de diseño ni resoluciones formales.	Adopción y consecuencias; débil independencia; participación <i>ex post</i> .
BA. Información de agencia.	Alto uso interno, opacidad externa; déficit de legitimidad.	X≈1.0 Y≈2.3 A≈0.5 B≈2.1	Unidad de evaluación del DIF estatal: informes internos para rediseñar atención temprana, no publicados.	Apertura e inteligibilidad; participación casi nula.
BB. Bien tipo club / caja negra.	Evidencia excluible y sin uso; captura y discrecionalidad.	X≈0.5 Y≈0.6 A≈0.3 B≈0.8	<i>Programa municipal de infraestructura escolar</i> : sin evaluaciones ni datos públicos.	Todas: no hay teoría de cambio, datos, ni reglas; alto riesgo para derechos.

Fuente: Elaboración propia.

Recomendaciones inmediatas (0–3 meses)	Recomendaciones estructurales (6–12 meses)	Indicadores de avance (seguimiento)
<ul style="list-style-type: none"> • Publicar paquetes replicables auditados (datos más diccionarios más código). • Acta de adopción por hallazgo («hallazgo–acción–responsable–plazo»). • Calendario de audiencias públicas. • Instrucción de resolución de adopción por evaluación (formato «cumple o explica»). • Comité metodológico para arbitraje y COI. • Talleres de validación pública (familias/escuelas). • Publicar inventario, cronograma y versiones públicas de informes. • Resúmenes en lectura fácil (cuidadores/niños, niñas y adolescentes). • Política FAIR mínima (metadatos, formatos abiertos). • Levantamiento de inventario de datos y evaluaciones. • Protocolo de evaluación con teoría de cambio. • Prueba de daño/interés público y versiones públicas. 	<ul style="list-style-type: none"> • Metaevaluación anual de calidad y uso. • Pilotos de protocolos registrados con universidades. • Comité ciudadano/técnico permanente. • Amarre normativo: vincular hallazgos a presupuesto/programación. • Pre-registro y publicación de negativos. • Tablero de seguimiento 3–6–12 meses. • Apertura de microdatos anonimizados con cinco seguros. • Capacitación en comunicación pública y visualización. • Canal de acceso seguro para academia/OSC. • Primeras evaluaciones piloto con preregistro. • Comité de transparencia y arbitraje externo. • Plan trienal de capacidad evaluativa (perfiles, formación, presupuesto). 	<ul style="list-style-type: none"> • Porcentaje estudios con revisión externa. • Tiempo para decisión desde publicación (días). • Porcentaje de recomendaciones implementadas. • Porcentaje de evaluaciones con resolución formal. • Número de ajustes de diseño presupuestal. • Porcentaje de evaluaciones con pre-registro/COI. • Porcentaje de bases de datos abiertos/FAIR. • Número de visitas/descargas y replicaciones externas. • Porcentaje de informes con lectura fácil. • Inventario y plan aprobados. • Número de evaluaciones piloto con preregistro. • Porcentaje de programas con teoría de cambio y línea base.

mienda probar distintas combinaciones de pesos y diferentes umbrales, documentar las decisiones tomadas y mostrar si los resultados se mantienen estables ante tales ajustes. Esta práctica, conocida como análisis de sensibilidad, permite transparentar los supuestos del modelo y fortalece su credibilidad ante la comunidad académica y los responsables de política.

3. La validez de la matriz depende de la calidad de los datos disponibles y de la consistencia con que se aplican los criterios de codificación. Como algunos indicadores, éstos requieren interpretación, por lo que es importante que la información sea revisada por más de una persona y se midan los niveles de acuerdo entre quienes la revisen. Esto ayuda a reducir el sesgo individual y mejora la confiabilidad de los resultados. También es necesario definir cómo se tratarán los datos incompletos o ambiguos y documentar las decisiones tomadas. Además, se debe considerar que la publicación de resultados o clasificaciones puede tener efectos no deseados; por ejemplo, incentivar que las instituciones busquen mejorar su posición sin cambiar de fondo sus prácticas de apertura y gobernanza. Para evitarlo es útil combinar indicadores menos manipulables, como auditorías externas o revisiones de código y datos, y mantener reglas claras de apertura y reproducibilidad.
4. La matriz no mide causalidad ni eficacia, sino capacidades institucionales para producir y usar evidencia con atributos de bien público. Su valor radica en mostrar patrones de apertura, gobernanza y participación, no en demostrar si una política es exitosa o no. Por eso sus resultados deben interpretarse más como una base para el diálogo y la mejora que como un juicio. Para reforzar su validez se recomienda contrastar las puntuaciones obtenidas con criterios externos, como el porcentaje de recomendaciones adoptadas, los tiempos de respuesta o las decisiones presupuestarias vinculadas a la evidencia. Asimismo, los análisis comparativos pueden complementarse con estudios de caso que ayuden a entender los procesos que hay detrás de las clasificaciones y con mediciones repetidas a lo largo del tiempo para identificar avances, retrocesos o aprendizajes institucionales.

Las estrategias de triangulación, verificación cruzada, mediciones repetidas y estudios cualitativos fortalecen la validez y confiabilidad de la matriz. Más que como una clasificación, la matriz funciona como diagnóstico y herramienta de aprendizaje para orientar reformas que vuelvan la evidencia un bien público, es decir, que sea accesible,

verificable y útil. En políticas de infancia, además, explicita obligaciones y salvaguardas de derechos —acceso, participación, privacidad y no daño— como criterios operativos de apertura y uso (United Nations, 1989; Comité de los Derechos del Niño, 2021; CPEUM, 2024; LGDNNA, 2024; LGTAIP, 2025; LGPDPSO, 2025).

Conclusiones

Este artículo parte de la tesis de que la evidencia no es en sí misma un bien público. Para que lo sea debe cumplir de manera simultánea con las condiciones de apertura usable (no exclusión e inteligibilidad), gobernanza para el uso (integridad, independencia y adopción con consecuencias), participación sustantiva y salvaguardas de derechos. Con base en ello, propusimos un marco conceptual y una matriz de gobernanza que traducen principios en criterios observables, y mostramos cómo esta arquitectura evita el «palomeo» y la transparencia ritual al orientar decisiones con trazabilidad pública. El aporte es doble: ofrece un marco analítico para identificar sistemáticamente cuándo las evaluaciones de programas sociales funcionan como mecanismos de rendición de cuentas y mejora, y argumenta a favor de transitar de una política basada en evidencia a la gobernanza de la evidencia.

En términos de rendición de cuentas, la evidencia contribuye como bien público cuando cierra el circuito de información-justificación-consecuencias. Esto ocurre si: 1) la apertura es usable y la mediación pública vuelve comprensibles los hallazgos; 2) la evidencia es verificable y contestable (trazabilidad de datos, métodos y decisiones); 3) su producción está protegida por integridad e independencia, y 4) existen rutas institucionales de respuesta que obligan a traducir hallazgos en actos verificables (decisiones, ajustes, reasignaciones, correcciones) bajo esquemas de «cumplir o explicar», con responsables, plazos y seguimiento. En políticas de infancia, además, la apertura debe ser responsable: privacidad, no daño y no revictimización no restringen el acceso; lo legitiman y vuelven socialmente utilizable, con lo que se evitan lecturas triunfalistas o instrumentalizaciones.

La matriz propuesta convierte la idea de «evidencia como bien público» en una herramienta práctica de diagnóstico. Permite ubicar sistemas en un mapa de apertura y gobernanza, identificar cuellos de botella y mover palancas para avanzar hacia el ideal de bien público robusto. Al asignar puntajes verificables a ejes y módulos, los sistemas pueden ver no sólo su nivel de apertura e integridad, sino también los déficits estructurales que limitan su valor público. En esa medida, la matriz no es una clasificación, sino

un instrumento de aprendizaje institucional que ayuda a priorizar mejoras y balancear la utilidad social con la protección de personas.

También reconocemos las limitaciones. La matriz no prueba causalidad ni mide impactos; clasifica capacidades. Su valor depende de que haya datos suficientes, reglas de codificación consistentes y se tomen decisiones de agregación transparentes. Por eso sugerimos prácticas de rigor sencillas: triangulación, doble codificación, mediciones repetidas y estudios de caso. De aquí en adelante, proponemos dos rutas. En investigación, poner a prueba las proposiciones P1–P5 con datos reales (uso social, legitimidad, pertinencia, inteligibilidad y confianza) y comparar trayectorias por estado y sector; en política pública, institucionalizar reglas simples y claras (prerregistro, manejo de conflictos de interés, publicación de resultados negativos, arbitraje externo y esquemas de «cumplir o explicar») para que la evidencia sea accesible, entendible, verificable y con consecuencias. Así, la evidencia deja de ser un bien reservado y se vuelve un bien público que fortalece la rendición de cuentas.

Referencias

- Aguilar-Villanueva, L. (2017). *Política pública*. Miguel Ángel Porrúa, Centro de Investigación y Docencia Económicas (CIDE).
- Arellano-Gault, D. y Ramírez-Macías, J. J. (2000). *Presupuesto dirigido a resultados: los dilemas de la evaluación por desempeño en contextos latinoamericanos. Lecciones desde México* [ponencia]. V Congreso Internacional de CLAD. Santo Domingo, República Dominicana. https://www.researchgate.net/publication/228360897_Presupuesto_dirigido_a_resultados_los_dilemas_de_la_evaluacion_por_desempeno_en_contextos_latinoamericanos_Lecciones_desde_Mexico
- Bovens, M. (2007). Analysing and assessing accountability: a conceptual framework. *European Law Journal*, 13(4), 447-468. <https://doi.org/10.1111/j.1468-0386.2007.00378.x>
- Cabrero, E. (2005). *Acción pública y desarrollo local*. Fondo de Cultura Económica.
- Cámara de Diputados del H. Congreso de la Unión (2024). *Constitución Política de los Estados Unidos Mexicanos*.
- Cámara de Diputados del H. Congreso de la Unión (2023). Ley del Sistema Nacional de Información Estadística y Geográfica (LSNIEG). *Diario Oficial de la Federación*. <https://www.diputados.gob.mx/LeyesBiblio/pdf/LSNIEG.pdf>

- Cámara de Diputados del H. Congreso de la Unión (2024). Ley General de los Derechos de Niñas, Niños y Adolescentes (LGDNNA). *Diario Oficial de la Federación*. <https://www.diputados.gob.mx/LeyesBiblio/pdf/LGDNNA.pdf>
- Cámara de Diputados del H. Congreso de la Unión (2025). Ley General de Protección de Datos Personales en Posesión de Sujetos Obligados (LGPDPPO). *Diario Oficial de la Federación*. <https://www.diputados.gob.mx/LeyesBiblio/pdf/LGPDPPO.pdf>
- Cámara de Diputados del H. Congreso de la Unión (2025). Ley General de Transparencia y Acceso a la Información Pública (LGTAIIP). *Diario Oficial de la Federación*. https://www.diputados.gob.mx/LeyesBiblio/ref/LGTAIIP/LGTAIIP_orig_20mar25.pdf
- Campbell, D. T. (1976). Assessing the impact of planned social change. [Occasional Paper Series, No. 8]. *Journal of MultiDisciplinary Evaluation (JMDE)*. <https://doi.org/10.56645/jmde.v7i15.297>
- Cejudo, G. M. (2013). Gobierno abierto en México: ¿etiqueta, principio o práctica? En Luna-Pla, I. y Bojórquez-Pereznieto, J. A. (coords.), *Gobierno abierto y el valor social de la información pública* (pp. 107-123). Universidad Nacional Autónoma de México (UNAM).
- Chambers, C. (2013). Registered reports: a new publishing initiative at Cortex. *Cortex*, 49(3), 609-610. <https://doi.org/10.1016/j.cortex.2012.12.016>
- Comité de los Derechos del Niño (2021, 2 de marzo). *CRC/C/GC/25: Observación general núm. 25 sobre los derechos de los niños en relación con el entorno digital*. Organización de las Naciones Unidas. <https://www.ohchr.org/es/documents/general-comments-and-recommendations/general-comment-no-25-2021-childrens-rights-relation>
- Consejo Nacional de Evaluación de la Política de Desarrollo Social (Coneval) (s.f.). *Evaluación de la política social. Evaluación de consistencia y resultados*. <https://www.coneval.org.mx/EvaluacionDS/Metodologia/Paginas/ECyR.aspx>
- Corte Interamericana de Derechos Humanos (CIDH) (2006, 19 de septiembre). *Caso Claude Reyes y otros vs. Chile*. [Serie 151, documentos de la CIDH]. https://corteidh.or.cr/docs/casos/articulos/seriec_151_esp.pdf?utm
- Desai, T., Ritchie, F. y Welpton, R. (2016). *Five safes: designing data access for research*. [Economics Working Paper Series 1601]. <https://www2.uwe.ac.uk/faculties/BBS/Documents/1601.pdf>

- Dwork, C. y Roth, A. (2017). The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3-4), 211-407. <https://www.cis.upenn.edu/~aaroht/Papers/privacybook.pdf>
- Espeland, W. y Sauder, M. (2007). Rankings and reactivity: how public measures recreate social worlds. *American Journal of Sociology*, 113(1), 1-40. <https://www.journals.uchicago.edu/doi/10.1086/517897>
- Fondo de las Naciones Unidas para la Infancia (UNICEF) (2025). *Responsible data for children*. UNICEF. <https://rd4c.org/es/resources?tab=tools>
- Fox, J. (2015). Social accountability: what does the evidence really say? *World Development*, 72, 346-361. <https://doi.org/10.1016/j.worlddev.2015.03.011>
- Franco, A., Malhotra, N. y Simonovits, G. (2014). Publication bias in the social sciences: unlocking the file drawer. *Science*, 345(6203), 1502-1505. <https://doi.org/10.1126/science.1255484>
- Fung, A., Graham, M. y Weil, D. (2007). *Full disclosure: the perils and promise of transparency*. Cambridge University Press. <https://doi.org/10.1017/9780521699617>
- Funnell, S. y Rogers, P. J. (2011). *Purposeful program theory: effective use of theories of change and logic models*. Jossey-Bass. <https://download.e-bookshelf.de/download/00/5819/65/L-G-0000581965-0002326787.pdf>
- Graizbord, D. (2024). *Indicators of democracy: the politics and promise of evaluation expertise in Mexico*. Stanford University Press.
- Henry, G. T. y Mark, M. M. (2003). Beyond use: understanding evaluation's influence on attitudes and actions. *American Journal of Evaluation*, 24(3), 293-314. [https://doi.org/10.1016/S1098-2140\(03\)00056-0](https://doi.org/10.1016/S1098-2140(03)00056-0)
- Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales (INAI) (2021). *Lineamientos generales para la clasificación y desclasificación de la información y la elaboración de versiones públicas*.
- Janssen, M., Charalabidis, Y. y Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. *Information Systems Management*, 29(4), 258-268. <https://doi.org/10.1080/10580530.2012.716740>
- Jasanoff, S. (ed.) (2004). *States of knowledge: the co-production of science and social order*. Routledge.
- Labin, S., Duffy, J., Meyers, D., Wandersman, A. y Lesesne, C. (2012). A research synthesis

- of the evaluation capacity building literature. *American Journal of Evaluation*, 33(3), 307-338. <https://doi.org/10.1177/1098214011434608>
- Merino, M., López Ayllón, S. y Cejudo, G. (eds.) (2010). *La estructura de la rendición de cuentas en México*. Instituto de Investigaciones Jurídicas (IIJ)-UNAM, CIDE.
- Murrieta, P., Caro, B. y Ruiz, M. (2011). *Evaluación del respeto a los derechos de la infancia*. Editorial Académica Española (EAE).
- Nosek, B. A. (2015). Promoting an open research culture. *Science*, 348(6242), 1422-1425. <https://doi.org/10.1126/science.aab2374>
- Nosek, A. B., Ebersole, C. R., DeHaven, A. C. y Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, 115(11), 2600-2606. <https://doi.org/10.1073/pnas.1708274114>
- Nutley, S., Walter, I. y Davies, H. (2007). *Using evidence: how research can inform public services*. Policy Press.
- Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) (2010). *Por buen camino. Guía para el seguimiento y la evaluación de proyectos comunitarios*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000186231_spa
- Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) (2021). *Recomendación de la UNESCO sobre la ciencia abierta*. https://unesdoc.unesco.org/ark:/48223/pf0000379949_spa
- Organisation for Economic Co-operation and Development (OECD) (2015). *Making open science a reality*. [Technology and Industry Policy Papers, 25.] OECD Science. https://www.oecd.org/en/publications/making-open-science-a-reality_5jrs2f963zs1-en.html
- Organisation for Economic Co-operation and Development (OECD) (2017). *Perspectivas del comportamiento y políticas públicas: lecciones de todo el mundo*. OECD. https://www.oecd.org/en/publications/behavioural-insights-and-public-policy_9789264270480-en.html
- Organisation for Economic Co-operation and Development (OECD) (2019). *Better criteria for better evaluation*. OECD. https://www.oecd.org/content/dam/oecd/en/publications/reports/2019/12/better-criteria-for-better-evaluation_f7a307eb/15a-9c26b-en.pdf
- Parkhurst, J. (2017). *The politics of evidence: from evidence-based policy to the good governance of evidence*. Routledge.

- Patton, M. Q. (2008). *Utilization-focused evaluation* (4a. ed.). Sage.
- Pawson, R. y Tilley, N. (1997). *Realistic evaluation*. Sage.
- Planas, A., Pineda-Herrero, P., Gil, E. y Sánchez, L. (2014). La metodología de la evaluación participativa de planes y acciones comunitarias. Tres experiencias de evaluación participativa en Catalunya. *Pedagogía Social*, (24), 105-134. <https://www.redalyc.org/pdf/1350/135043652006.pdf>
- Porter, T. M. (1995). *Trust in numbers: the pursuit of objectivity in science and public life*. Princeton University Press.
- Power, M. (1997). *The audit society: rituals of verification*. Oxford University Press.
- Preskill, H. y Boyle, S. (2008). A multidisciplinary model of evaluation capacity building. *American Journal of Evaluation*, 29(4), 443-459. <https://doi.org/10.1177/1098214008324182>
- Rosenthal, R. (1979). The «file drawer problem» and tolerance for null results. *Psychological Bulletin*, 86(3), 648-641. <https://pages.ucsd.edu/~cmckenzie/Rosenthal1979PsychBulletin.pdf>
- Rossi, P. H., Lipsey, M. y Freeman, H. E. (2004). *Evaluation: a systematic approach* (7a. ed.). Sage.
- Secretaría de Hacienda y Crédito Público (SHCP) (2016). *Guía para el diseño de la matriz de indicadores para resultados*. SHCP.
- Secretaría de la Suprema Corte de Justicia de la Nación (2024, 20 de diciembre). *Artículo 6º constitucional: Libertad de expresión y derecho de acceso a la información*. https://www.scjn.gob.mx/sites/default/files/marco_normativo/documeto/2025-01/Articulo_6_Constitucional_20241220.pdf
- Silva, C. (2024). Presupuesto basado en resultados y sistema de evaluación del desempeño en México: evolución, limitantes y aprendizajes. *Economía UNAM*, (447), 3-30.
- Sistema Nacional de Información Estadística y Geográfica (SNIEG) (2021, 29 de octubre). Política para la gestión de la confidencialidad en la información estadística y geográfica. *Diario Oficial de la Federación*. https://www.snieg.mx/Documentos/Normatividad/Vigente/politica_inf_estad_geog.pdf
- Stiglitz, J. E. (1999). Knowledge as a global public good. En I. Kaul, I. Grunberg y M. A. Stern (eds.), *Global public goods: international cooperation in the 21st century*. Oxford University Press, United Nations Development Programme (UNDP).

- Stufflebeam, D. y Coryn, C. (2014). *Evaluation theory, models, and applications*. Jossey-Bass.
- United Nations (1989). *Convention on the Rights of the Child*. United Nations. <https://treaties.un.org/doc/Publication/MTDSG/Volume%20I/Chapter%20IV/IV-11.en.pdf>
- Wholey, J. S. (2010). Exploratory evaluation. En J. S. Wholey, H. P. Hatry y K. E. Newcomer (eds.), *Handbook of practical program evaluation* (pp. 81-99). Jossey-Bass.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Da Silva-Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Sci Data*, 3(160018). <https://doi.org/10.1038/sdata.2016.18>
- Zuiderwijk, A. y Janssen, M. (2014). Open data policies, their implementation and impact: a framework for comparison. *Government Information Quarterly*, 31(1), 17-29. <https://doi.org/10.1016/j.giq.2013.04.003>