Lineamientos para la Gobernanza de Datos

buenosaires.gob.ar/datosabiertos

Buenos Aires Ciudad

Vamos por más

ÍNDICE

1. INTRODUCCIÓN

Los datos se han convertido en uno de los activos más valiosos con los que cuentan las organizaciones competitivas en el mundo. Su uso oportuno permite tomar decisiones basadas en evidencia, identificar tendencias, optimizar recursos y anticipar problemas. Contar con datos precisos y de alta calidad es fundamental para mantenerse competitivo y responder eficazmente a los desafíos y oportunidades de diversos sectores.

En este nuevo contexto, la Secretaría de Innovación y Transformación Digital del Gobierno de la Ciudad de Buenos Aires (en adelante, SECITD) está trabajando en el proyecto de **Lineamientos para la Gobernanza de Datos**, con el objetivo de profundizar y sentar las bases para la implementación de una política integral en materia de gestión de datos, abarcando tanto su obtención, generación, protección, uso, análisis, integración y almacenamiento, con el objetivo de tomar decisiones basadas en evidencia y mejorar la vida de los vecinos y vecinas que viven y transitan por la Ciudad.

El proyecto nace como consecuencia del análisis realizado sobre la capacidad del gobierno para convertirse en una organización eficiente y competitiva en un ecosistema cada vez más regido por los datos. Dado que las distintas dependencias gubernamentales, tienen entre sus atribuciones el recopilar, procesar y analizar información clave para la toma de decisiones y la formulación de políticas, para que esta información sea interoperable y pueda ser utilizada eficazmente por las diversas dependencias, es indispensable establecer pautas comunes que garanticen su consistencia, comparabilidad y usabilidad.

En esta materia, el Gobierno de la Ciudad Autónoma de Buenos Aires (en adelante, GCABA) debe:

1. **Generar una cultura de uso de datos** entre líderes, funcionarios y equipos técnicos.
2. **Promover la mejora** de competencias y capacidades del capital humano.
3. **Enmarcar roles y responsabilidades** de los actores intervinientes en la implementación de la política de datos.
4. **Mejorar la fase de generación** de estándares que faciliten la interoperabilidad de los datos.
5. **Facilitar los espacios** de cooperación y aprendizaje entre las áreas.

Los siguientes **Lineamientos para la Gobernanza de Datos** buscan establecer una ruta de trabajo que permita la mejora continua y la adaptabilidad del gobierno, transformándose en una organización data-driven que gobierne, gestione y democratice los datos de manera

estratégica para ser más ágiles y efectivos al momento de tomar decisiones que impacten positivamente en la calidad de vida de los vecinos y vecinas.

1.1. ¿Por qué lineamientos?

En la actualidad, la mayoría de las áreas acostumbran a trabajar solo con **información generada internamente.** Muchas veces, por **desconocimiento de los activos disponibles** y por la **ausencia de criterios claros,** se **duplican los datos** y se dificulta la integración.

Tener lineamientos es contar con una **ruta clara de cómo gobernar los datos** para implementar políticas, procedimientos y normas que promuevan una toma de decisiones más eficaz.

Establecer los Lineamientos para la Gobernanza de Datos es poder documentar y guiar todo el proceso, proporcionando un **instructivo necesario para trabajar con los datos,** ya que eliminan las múltiples interpretaciones, fomentan la cooperación entre las áreas, institucionalizan y protegen el flujo de datos del Gobierno y de los vecinos y vecinas, y formalizan procesos.

La importancia de los lineamientos se basa en que permiten **establecer criterios claros y únicos,** e identificar objetivos que mejoren el modo en que gestionamos y compartimos los datos, de manera ágil y segura.

1.2. ¿Qué pasa cuando no gobernamos los datos?

Para entender qué es lo que sucede cuando no gobernamos los datos, planteamos la siguiente situación a modo de ejemplo:

Le encomiendan un proyecto a una persona, en el cual tiene 6 meses para decidir qué calles de la Ciudad de Buenos Aires se pueden peatonalizar (teniendo en cuenta que, durante el contexto de pandemia por COVID-19, transformar el espacio público se volvió fundamental). Para desarrollar el proyecto y presentar los resultados, esta persona necesita información sobre cuánta gente transita por las calles; cuántos autos, motos y colectivos circulan por cada calle y qué comercios funcionan en las calles.

Lo primero que se preguntaría la persona a cargo del proyecto sería lo siguiente: ¿de dónde obtener esos datos? ¿Existen estos datos? ¿Dónde se encuentran alojados? ¿Tendrán datos históricos? ¿Hay algún área que cuente con datos que complementen el análisis? ¿Son datos de calidad?

Suponiendo que dichos datos existen, es importante que se encuentren en forma ordenada, que se puedan usar, reutilizar y que exista un intercambio de datos entre áreas para un buen uso de los mismos.

Es necesario pensar en qué pasaría si no están estas condiciones. Si pasan dos meses y la persona sigue sin poder resolver de dónde sacar la información necesaria, será complejo continuar con el proyecto asignado.

Esto ocurre cuando **no hay una buena gobernanza de datos:** puede generar retrasos y demoras significativas hasta que se logra obtener los datos, tomar decisiones, y concretar proyectos.

2. CLASIFICACIÓN DE DATOS

2.1. Definición e importancia de la clasificación de datos

La clasificación de datos es el proceso por el cual se organizan los datos que posee cualquier organización o área en categorías, con el objetivo de facilitar su tratamiento, en términos de acceso, intercambio y reutilización, en función de sus particularidades y normativas específicas.

Este proceso permite realizar un tratamiento más eficiente, seguro y lícito de los datos, garantizando la protección de los derechos de todas las personas.

Una clasificación sólida permite reducir riesgos y definir responsabilidades en la gestión de los datos, otorgando **seguridad, confidencialidad, integridad, y privacidad** a **los mismos.**

Este apartado busca ofrecer **criterios indicadores y orientadores** en el ámbito público del Gobierno de la Ciudad Autónoma de Buenos Aires sobre las **diferentes categorías de datos,** en consonancia con la Ley N° 1845 de Protección de Datos Personales de la Ciudad Autónoma de Buenos Aires y la normativa asociada, contribuyendo así a la **adopción de políticas de privacidad y seguridad.**

Es esencial resaltar que la clasificación de datos varía significativamente según el contexto en el que se aplique. Por lo tanto, y más allá de seguir las directrices de esta guía, es crucial realizar un análisis detallado de cada situación particular.

2.2. Clasificación de datos

2.2.1. Datos públicos

Se entiende por **datos públicos** a todo dato, información, constancia o documento, cualquiera sea el soporte en el que esté contenido o representado, que hubiere sido o debiera ser creado u obtenido por el Estado, o que obrare o debiera obrar en su poder o bajo su control, para el cumplimiento de sus misiones y funciones.

Es decir, podríamos considerar dato público a toda aquella información contenida en documentos escritos, fotográficos, grabaciones, soporte magnético, digital o en cualquier otro formato; incluyendo bases de datos, acuerdos, directivas, reportes, estudios, oficios, proyectos de ley, disposiciones, resoluciones, providencias, expedientes, informes, actas, circulares, contratos, convenios, estadísticas,

instructivos, dictámenes, boletines o cualquier otra información registrada en cualquier fecha, forma y soporte; que haya sido creada u obtenida por el órgano requerido, y que se encuentre en su posesión y bajo su control.

Es fundamental distinguir que **no todos los datos públicos son de acceso público**, ya que los mismos pueden estar sujetos a regulaciones que limitan su difusión por motivos de seguridad, privacidad o protección de datos como veremos más adelante.

Por ello, no todos los datos públicos se encuentran publicados ni se consideran datos abiertos. Es decir, dentro de esta categoría se creó una forma de compartir diferentes datos que genera el Estado en formatos abiertos que facilitan su utilización para todos los ciudadanos. La característica principal de estos **datos públicos abiertos** es que **cualquier persona puede acceder, usar y compartir libremente los mismos.**

A modo de ejemplo, se considera información pública:

- El presupuesto otorgado a cada área.
- Las contrataciones públicas.
- Los informes de auditorías.
- Los mecanismos directos de presentación de denuncias.
- Datos que permitan la participación de la ciudadanía en la toma de decisiones.
- Cantidad de expedientes que tramita el GCABA por año.

Los datos que se comparten a través de esos procesos son llamados datos abiertos. En el Gobierno de la Ciudad de Buenos Aires, la plataforma oficial en donde se presenta esta información es Buenos Aires Data.

2.2.2. Datos personales

Los datos personales permiten identificar a personas físicas o de existencia ideal, directa o indirectamente, mediante uno o varios elementos característicos de su identidad.

En la Ciudad Autónoma de Buenos Aires los mismos se encuentran regulados por la Ley N° 1.845 de Protección de Datos Personales, que los define como *"información de cualquier tipo referida a personas físicas o de existencia ideal, determinadas o determinables"* y que tiene por objeto regular el tratamiento de estos datos personales asentados o destinados a ser asentados en archivos, registros, bases o bancos de datos del sector público de la Ciudad de Buenos Aires, a los fines de garantizar el

derecho al honor, a la intimidad y a la autodeterminación informativa, de conformidad a lo establecido por el artículo 16 de la Constitución de la Ciudad de Buenos Aires.

La referencia que realiza la ley, al carácter determinable de las personas, implica que, si bien los datos pueden no identificar a la persona de forma unívoca, lo cierto es que por medio de un procedimiento o uniendo diversas fuentes de datos se la podría identificar.

En función de lo expuesto, los datos personales son, por ejemplo: nombre, Documento Nacional de Identidad (DNI), domicilio, edad, fecha de nacimiento, número telefónico, correo electrónico personal, CUIT, características físicas, genéticas o biométricas; patrimonio, trayectoria académica, laboral o profesional.

Asimismo, no son considerados datos personales aquellos que no hagan referencia a una persona, por ejemplo, una dirección de correo electrónico del tipo **info@empresa.com**, datos anonimizados y/o disociados, siempre y cuando el proceso de desambiguación del dato no sea reversible.

A continuación, se detallan los distintos tipos de datos personales:

2.2.2.1. Datos personales de carácter sensible

Los datos sensibles son aquellos datos personales que refieren a características físicas o morales de las personas o a hechos o circunstancias de su vida privada o intimidad, como pueden ser aquellos datos que revelan origen racial o étnico, opiniones políticas, convicciones religiosas o morales, afiliación sindical, información referente a la salud o a la vida sexual o cualquier otro **dato que pueda producir, por su naturaleza o su contexto, algún trato discriminatorio a su titular**.

En esta categoría resulta relevante efectuar una aclaración, con el objeto de resaltar la importancia del contexto y la finalidad en la utilización o divulgación del dato o información.

La Ley N° 1.845 de Protección de Datos Personales no incluye dentro de los datos sensibles al domicilio. No obstante, debemos poner especial atención sobre este punto debido a que el mismo **en algunas circunstancias puede indicar un dato sensible.** Así, por ejemplo, en el supuesto de casos de víctimas de violencia de género en donde la reserva del domicilio actual de la víctima resulta esencial, dicho dato adquiere carácter de sensible en función de ese contexto particular. En estos casos, su posible divulgación sugiere un

riesgo para la persona.

2.2.2.2. Datos de salud

Los datos de salud son aquellos que se refieren a la condición física y mental de una persona, incluyendo diagnósticos, tratamientos, historial médico, aspectos psicológicos, entre otros.

Los datos de salud son fundamentales, entre otras cuestiones, para brindar una atención médica personalizada, fomentar la investigación y desarrollo, y facilitar la prevención y detección de ciertas condiciones médicas.

Dada la sensibilidad de esta información, los datos de salud están sujetos a regulaciones de protección de datos, y será crucial el contexto para determinar su tratamiento.

A modo de ejemplo, los establecimientos de salud públicos o privados y los profesionales vinculados a los mismos podrán recolectar y tratar los datos personales relativos a la salud física o mental de los pacientes que acudan a los mismos o que estén o hubieren estado bajo tratamiento de aquellos, respetando los principios del secreto profesional a fin de ejercer sus funciones, previo consentimiento de su titular. No obstante, se exime de requerir el consentimiento cuando, por ejemplo, existiera un grave peligro para la salud pública o una situación de emergencia.

2.2.2.3. Datos fiscales

Los datos fiscales son todos aquellos relacionados con la situación fiscal de las personas y entidades.

Respecto a este tipo de datos, en el derecho argentino y específicamente en la Ley N° 6.505 y sus modificaciones, texto ordenado por Decreto 174/24 se establece el secreto fiscal.

El mismo refiere a que todas las declaraciones juradas, comunicaciones e informes que presentan los contribuyentes, responsables o terceros en cumplimiento de las obligaciones establecidas por el Código Fiscal son de carácter secreto.

No obstante, no se aplica el secreto fiscal en los casos de solicitudes de información que contengan datos de carácter administrativo, como apellido y

nombres, denominación o razón social, domicilio, código postal; por organismos fiscales nacionales, provinciales y municipales, a condición de reciprocidad; por personas, empresas o entidades a quienes la Administración Gubernamental de Ingresos Públicos les encomiende la realización de tareas administrativas, relevamientos de estadísticas, computación, procesamiento de información, confección de padrones y otras necesarias para el cumplimiento de sus fines; por organismos de seguridad social de las diferentes jurisdicciones estatales; a solicitud mediante oficio judicial; por la Unidad de Información Financiera (UIF); y/o sea solicitada por el Ministerio Público Fiscal y unidades específicas que lo integren.

2.2.2.4. Datos antecedentes penales

Los datos relativos a antecedentes penales o contravencionales solo pueden ser objeto de tratamiento por parte de las autoridades públicas competentes, en el marco de las leyes y reglamentaciones respectivas.

3. TRANSFERENCIA DE DATOS

3.1. Definición e importancia de la transferencia de datos

La **transferencia de datos** se refiere al proceso mediante el cual la información es transmitida de un punto a otro, ya sea dentro de una misma organización o entre diferentes entidades. Utilizamos el término "transferencia" como un concepto general que engloba diversas formas de compartir datos, tales como la cesión, la interoperabilidad, o la transmisión para almacenamiento y resguardo. Este proceso puede realizarse a través de canales de comunicación tanto físicos como digitales, los cuales deben estar debidamente asegurados para proteger la integridad y confidencialidad de los datos durante su tránsito.

En un entorno donde la información fluye continuamente, asegurar la transferencia de datos es fundamental para prevenir accesos no autorizados, evitar la interceptación malintencionada y garantizar que los datos lleguen a su destino sin alteraciones. La falta de medidas adecuadas en la transferencia de datos puede dar lugar a brechas de seguridad, pérdida de información crítica y, en última instancia, afectar la confianza y la reputación de las organizaciones.

Por ello, es crucial establecer lineamientos claros que guíen el **proceso de transferencia de datos**, asegurando que se realice de manera segura, cumpliendo con las normativas vigentes y utilizando las **mejores prácticas de seguridad**.

3.2. Transferencia de datos en función de su clasificación

De acuerdo con la clasificación de datos previamente establecida, es fundamental conocer los requisitos administrativos necesarios para compartir información. A continuación, se describen los requisitos que deben cumplirse para garantizar una adecuada protección y cumplimiento normativo.

3.2.1. Datos públicos

Cuando los datos refieran a información pública es de aplicación la Ley N° 104 de Acceso a la Información Pública. Según lo que indica, **toda persona tendrá derecho a solicitar y a recibir información completa, veraz, adecuada y oportuna.** No será necesario acreditar derecho subjetivo, interés legítimo ni razones que motiven solicitar datos públicos.

Es así que los sujetos obligados, por ejemplo, todos los órganos pertenecientes a la administración central, deberán proveer la información contenida en documentos escritos, fotográficos, grabaciones, soporte magnético, digital o en cualquier otro formato; incluyendo bases de datos, acuerdos, directivas, reportes, estudios, oficios, proyectos de ley, disposiciones, resoluciones, providencias, expedientes, informes, actas, circulares, contratos, convenios, estadísticas, instructivos, dictámenes, boletines o cualquier otra información registrada en cualquier fecha, forma y soporte; que haya sido creada u obtenida por el órgano requerido, y que se encuentre en su posesión y bajo su control.

Es importante destacar que **el acceso a la información pública es gratuito**, salvo en los casos en que se necesite su reproducción, cuyos costos serán asumidos por el solicitante.

Los sujetos obligados bajo la normativa deben proporcionar la información solicitada y deben realizar todas las acciones necesarias para hacerlo. Las únicas excepciones a este principio general son las siguientes:

- Información que pueda afectar la intimidad de las personas o refiera a datos sensibles, a menos que se puedan aplicar mecanismos de disociación, se cuente con el consentimiento expreso, o no sea necesario el consentimiento.
- Información protegida por derechos de autor, secreto profesional, industrial o comercial, que pudiera afectar la competitividad o lesionar intereses del sujeto obligado.
- Información cuya publicidad pudiera revelar estrategias judiciales, técnicas o procedimientos de investigación, salvo que existan mecanismos de disociación.
- Información de terceros obtenida en carácter confidencial, que pudiera poner en peligro el correcto funcionamiento del sistema financiero, bancario o estadístico o que esté protegida por el secreto bancario, fiscal o estadístico.
- Información cuya divulgación pudiera poner en riesgo la seguridad pública o que sea información de carácter judicial que esté protegida por compromisos internacionales asumidos por la Ciudad Autónoma de Buenos Aires.

- Información contenida en notas internas u opiniones producidas como parte del proceso previo a la toma de decisión de autoridad pública que no formen parte de los expedientes.

El tratamiento de los datos públicos se alinea, también, con el Plan de Transparencia Activa.

El Gobierno de la Ciudad de Buenos Aires incentiva la publicación de la mayor cantidad de datos públicos posibles de manera completa y actualizada en el Portal de Datos Abiertos de la Ciudad, Buenos Aires Data, promoviendo de esta manera la transparencia y el acceso a la información.

3.2.2. Datos personales

El tratamiento de **datos personales** se rige por el principio general de que, para su transferencia, se requiere el consentimiento libre, expreso e informado de la persona titular, el que deberá constar por escrito, o por otro medio que permita se le equipare, de acuerdo a las circunstancias

Asimismo, para el supuesto de datos personales recolectados a través de internet, se considerará cumplido el requisito del **consentimiento** previo cuando el titular del dato acepte la **política de privacidad** publicada en el sitio de Internet a través del cual se recolecten sus datos.

No obstante, el artículo N° 7 de la Ley N° 1.845 de Protección de Datos Personales establece excepciones al requisito del consentimiento para el tratamiento de los datos personales. La ley establece que **no será necesario el consentimiento** cuando:

- Los datos personales se recaben **para el ejercicio de funciones propias de los poderes** del Gobierno de la Ciudad de Buenos Aires, o en virtud de una obligación legal.
- Los datos personales **se obtengan de fuentes de acceso público irrestricto**.
- Se trate de datos personales relativos a la salud de las personas y su tratamiento sea necesario por **razones de salud pública y emergencia** establecidas por autoridad competente y debidamente fundadas.
- Se trate de **listados** cuyos datos se limiten a nombre, documento nacional de identidad, identificación tributaria o previsional, ocupación, fecha de nacimiento y domicilio.

- Cuando la cesión **"se realice entre órganos del sector público de la Ciudad de Buenos Aires en forma directa, siempre que sea en cumplimiento de sus respectivas competencias"**. Esta excepción se extiende a otros organismos o entidades externas al Gobierno de la Ciudad, siempre que puedan declarar un interés legítimo o competencia que justifique la solicitud de los datos.

Este aspecto se desarrollará en mayor detalle en la siguiente sección, donde se abordará la **transferencia de datos en función del solicitante**.

3.2.2.1. Datos sensibles

Es necesario obtener el **consentimiento expreso, libre e informado de la persona titular antes de compartirlos**, salvo que existan razones de interés general fundadas por ley o un requerimiento judicial, o cuando sean tratados con finalidades estadísticas o científicas, siempre y cuando no puedan ser identificados sus titulares.

Para datos de salud, la normativa establece excepciones al consentimiento en caso de que:

- La cesión se realice entre organismos sanitarios, tanto a nivel nacional como provincial y de la Ciudad Autónoma de Buenos Aires, con fines compatibles con el tratamiento médico aplicable a las personas titulares de los datos, en función de las competencias explícitas e implícitas que les hayan sido conferidas por ley.
- Lo establezca expresamente una ley especial referida a cuestiones sensibles, en particular sobre salud pública, emergencias y seguridad.
- La cesión sea necesaria por razones de salud pública, de emergencia o para la realización de estudios epidemiológicos, en tanto se preserve la identidad de los titulares de los datos mediante mecanismos de disociación adecuados.

Un ejemplo de interés general fundado en ley para el tratamiento de datos sensibles es el manejo de datos personales referidos a la salud y a la geolocalización de los titulares durante la pandemia por COVID-19.

3.2.3. Datos anonimizados

Aunque los datos anonimizados no constituyen una categoría de dato por sí mismos, es relevante destacar que, conforme a lo indicado en algunas excepciones, pueden ser compartidos cuando han sido disociados de la persona a la que se refieren. Esto se aplica, por ejemplo, en fines estadísticos, para ampliar el acceso a la información pública o, en ciertos casos, para rendición de cuentas.

Dado que estos datos ya no están vinculados a una persona identificable, no requieren una solicitud previa para su transferencia, ya que, al haber sido anonimizada la información personal, no se vulneran los principios de protección de datos.

3.3. Transferencia de datos en función del solicitante

La transferencia de datos, en función del solicitante, debe considerar la clasificación de los datos previamente establecida y el perfil de quien solicita el acceso a la información. Esta relación determina los requisitos de autorización, formalidad y protección necesarios para asegurar una transferencia conforme a las normativas vigentes.

Es importante destacar que, dado que los datos públicos ya han sido abordados previamente y no requieren requisitos específicos para su transferencia, esta sección se centrará exclusivamente en la transmisión de datos que sí requieren consideraciones adicionales, como los datos personales y sensibles.

3.3.1. Órganos del sector público de la Ciudad Autónoma de Buenos Aires

Las distintas dependencias del ámbito público, según los artículos 7 y 10 de la Ley N° 1.845 de Protección de Datos Personales, podrán transferir entre sí **datos personales**, de forma directa y sin el consentimiento del titular de los datos, en la medida en que sea necesario para el cumplimiento de sus respectivas competencias o en virtud de una obligación legal, y respeten que:

1. El cedente haya obtenido los datos en ejercicio de sus funciones.
2. El cesionario utilice los datos pretendidos para una finalidad que se encuentre dentro del marco de su competencia.

3. Los datos involucrados sean adecuados y no excedan el límite de lo necesario en relación con esta última finalidad.

Se deberá efectuar a través de un medio informático auditable que permita acceder a la finalidad de la transferencia, las competencias y funciones de cada parte, el tipo de dato y su fuente. Las consideraciones específicas sobre este proceso se tratarán en la siguiente sección, dado que dependen de la finalidad y el tratamiento que se pretenda dar a los datos transferidos.

Por otro lado, en caso de que los datos personales a transferir sean **sensibles**, conforme lo establecido en la Ley N° 1.845 de Protección de Datos Personales, deberá existir consentimiento previo, libre, expreso e informado de la persona titular de los datos, a la que se le debe informar sobre la finalidad de la transferencia e identificar al cesionario o los elementos que permitan hacerlo, salvo que existan razones de interés general fundadas por ley.

3.3.2. Entidades privadas

Los organismos públicos podrán **transferir datos personales no sensibles** al sector privado **de manera no masiva**, siempre que dicha transferencia se realice en el marco de sus funciones institucionales o en cumplimiento de una obligación legal. En estos casos, es necesario obtener el consentimiento previo, libre y expreso del titular de los datos, salvo cuando la transferencia esté justificada bajo el principio del interés legítimo, es decir, cuando los beneficios de la transferencia superen cualquier potencial afectación a los derechos individuales.

Es fundamental realizar un análisis cuidadoso que equilibre el derecho a la privacidad del titular con el interés legítimo del solicitante. Este análisis determinará si es indispensable solicitar el consentimiento explícito del titular para la transferencia o si, en casos justificados, puede omitirse bajo el principio del interés legítimo.

Es fundamental que el organismo emisor del dato garantice que la transferencia no perjudique a los titulares, priorizando el respeto a los derechos personales. Además, debe asegurarse de que la entidad receptora cumpla con las normativas y principios de protección de datos, verificando que su uso sea seguro y conforme a las leyes aplicables.

Por otro lado, para la cesión de los **datos sensibles al sector privado**, será

necesario contar con el **consentimiento de la persona titular del dato** de manera previa, salvo **requerimiento judicial o una ley que así lo disponga.**

No obstante, la normativa establece excepciones al consentimiento para la **transferencia de datos de salud al sector privado**, en los siguientes casos:

- Si así lo establece expresamente una ley especial referida a cuestiones sensibles, en particular sobre salud pública, emergencias y seguridad.
- Si la cesión de los mismos es necesaria por razones de salud pública, de emergencia o para la realización de estudios epidemiológicos, en tanto se preserve la identidad de los titulares de los datos mediante mecanismos de disociación adecuados, conforme lo establece el artículo N° 11 inciso D de la Ley Nacional N° 25.326 de Protección de Datos Personales.

3.3.2.1 Personas físicas

La transferencia de datos también podrá realizarse a toda persona que así lo solicite en el marco y con el alcance y limitaciones que establece la Ley N° 104 de Acceso a la Información Pública.

Los vecinos y vecinas podrán **solicitar y recibir información completa, veraz, adecuada y oportuna** para ejercer el **derecho de acceso a la información pública**, sin necesidad de acreditar derecho subjetivo, interés legítimo o razones que motiven la petición.

Sin embargo, este derecho tendrá las limitaciones establecidas por la ley, pudiendo los sujetos obligados exceptuarse de proveer la información solicitada cuando se configure alguno de los supuestos allí establecidos, como por ejemplo, que afecte la intimidad de las personas o se trate de información referida a datos sensibles.

3.3.3. Otras jurisdicciones

3.3.3.1. Nacional

La transferencia de datos interprovincial o municipal entre organismos públicos de distinta jurisdicción podrá realizarse siempre y cuando la provincia o municipio proporcione un nivel adecuado de protección, entendiendo por tal al menos estar adherido a la normativa sobre protección de datos personales que se encuentre

vigente en la República Argentina al momento de la transferencia.

Asimismo, podrá transferirse datos en cualquiera de los siguientes supuestos:

- Colaboración judicial interjurisdiccional.
- Intercambio de datos de carácter médico para el tratamiento del afectado o investigación epidemiológica.
- Información referida a transferencias bancarias, en lo relativo a las transacciones respectivas y conforme la legislación que les resulte aplicable.
- A requerimiento de la autoridad judicial y en el marco de una causa.
- Lucha contra el crimen organizado, el terrorismo y el narcotráfico, y se realice a requerimiento de la autoridad judicial y en el marco de una causa
- Exista consentimiento del titular de los datos

3.3.3.2. Internacional

En lo que respecta a la transferencia internacional de datos personales, si bien la Ley N° 1.845 de Protección de Datos Personales señala algunos principios generales, este tipo tratamiento es legislado en la Ley Nacional N° 25.326 de Protección de Datos Personales, regulado en su Decreto Reglamentario N° 1.558 de 2001 y por la Disposición N° 60 de 2016, publicada por la Dirección Nacional de Protección de Datos Personales, toda vez que se trata de una acción internacional que trasciende las fronteras.

La Red Iberoamericana de Protección de Datos Personales, de la cual la Secretaría de Innovación y Transformación Digital del Gobierno de la Ciudad de Buenos Aires forma parte como miembro observador, establece:

Asimismo, entendiendo el interés que posee esta transferencia, la Red mencionada ha elaborado, como marco de recomendación para los Estados, los Estándares de Protección de Datos.

En el plano del derecho internacional, el Convenio N° 108 "Para la protección de las personas con respecto al tratamiento automatizado de datos de carácter personal", del que por cierto, Argentina es parte desde el primero de junio de 2019; propone como principio general que los flujos transfronterizos de datos de carácter personal, con destino al territorio de otro Estado Parte, ***no se prohíban o sometan a una autorización especial.***

Este presupuesto tiene sus excepciones, las que encontramos en el **artículo N°12** del **Convenio**:

- Cuando la legislación de una de las partes prohíba la transferencia internacional de determinados datos personales, a menos que la reglamentación de la otra establezca una protección equivalente. Este caso es el que encontramos en la legislación argentina, referido en la Disposición N° 60/16 mencionada precedentemente.

- Cuando la transferencia se realice hacia un Estado no Parte del convenio, por intermedio de un Estado que sí es parte. Esta prohibición se estableció a efectos de evitar que la transferencia tenga como resultado burlar la legislación del Estado no parte.

En lo que respecta a la normativa nacional en la materia, los requisitos a cumplimentar para la transferencia internacional de datos personales dependen si el país de destino es considerado "país adecuado" o no, consideración que se aclarará en el próximo apartado.

3.3.3.2.1. Transferencia internacional a países "adecuados"

Según la Ley Nacional N° 25.326 de Protección de Datos Personales, los "países adecuados" son aquellos que cuentan con una normativa vigente similar a la existente en la República en materia de protección y seguridad de datos personales.

La calidad de "adecuado" puede corresponder al sector público, privado o ambos.

En la actualidad, la Disposición N° 60 es la que establece qué países y sectores son considerados "adecuados", a saber: *"los estados miembros de la Unión Europea y miembros del espacio económico europeo (EEE), Reino Unido de Gran Bretaña e Irlanda del Norte, Confederación Suiza, Guernsey, Jersey, Isla de Man, Islas Feroe, Canadá* ***sólo respecto de su sector privado****, Principado de Andorra, Nueva Zelanda, República Oriental del Uruguay y Estado de Israel* ***sólo respecto de los datos que reciban un tratamiento automatizado*****".**

Asimismo, la Ley N° 1.845 de Protección de Datos Personales y su Decreto Reglamentario presume que un Estado u organismo internacional o supranacional proporciona un nivel adecuado de protección cuando dicha

tutela se deriva directamente del ordenamiento jurídico vigente, o de sistemas de autorregulación, o del amparo que establezcan las cláusulas contractuales que prevean la protección de datos personales.

3.3.3.2.2. Transferencia internacional a países considerados "no adecuados"

Aquellas responsables de tratamiento que efectúen transferencias de datos personales a países que no posean legislación adecuada en los términos del artículo N° 12 de la Ley N° 25.326 y su Decreto Reglamentario N° 1.558/01 deberán:

a. Suscribir contratos con las cláusulas contractuales "tipo de transferencia internacional para la transferencia y prestación de servicios", incorporadas en los Anexos I y II de la Disposición 60/2016, a fin de garantizar un nivel adecuado de protección de datos personales en los términos del artículo N° 12 de la Ley N° 25.326 y del Anexo I al Decreto N° 1.558/01.

b. En caso que las responsables del tratamiento de datos utilicen contratos que difieran de los modelos aprobados en los Anexos I y II de la Disposición 60/2016 o no contengan los principios, garantías y contenidos relativos a la protección de los datos personales previstos en los modelos aprobados, deberán solicitar su aprobación ante la Dirección Nacional de Protección de Datos Personales, dentro de los treinta días corridos de su firma.

3.4. Transferencia de datos en función de la finalidad

Además de considerar la clasificación del dato y el solicitante, es fundamental adaptar el proceso de transferencia de datos a la finalidad específica para la cual se requiere la información. La finalidad de la transferencia determina no solo la frecuencia y el método utilizado, sino también las medidas de seguridad, el nivel de automatización y los requisitos técnicos necesarios para asegurar una transferencia efectiva y segura.

Es esencial que, para obtener la información, se recurra siempre a la fuente auténtica, es decir, la entidad que registra, resguarda, mantiene y emite el dato. En caso de que no sea posible acceder a la fuente auténtica, se puede acudir a una fuente secundaria autorizada, como por ejemplo, la Administración Gubernamental de Ingresos Públicos (AGIP) para datos

fiscales provenientes de la Administración Federal de Ingresos Públicos (AFIP), o la plataforma miBA de la Ciudad de Buenos Aires para identidad digital que obtiene datos del Registro Nacional de las Personas (ReNaPer).

El Gobierno de la Ciudad Autónoma de Buenos Aires ha establecido procedimientos para que estos procesos de envío de información sean eficientes, seguros y transparentes, respetando las normativas vigentes y garantizando el control adecuado sobre el acceso y uso de la información.

A continuación, se detallan los lineamientos que deben considerarse para garantizar una transferencia adecuada de datos en función de la finalidad.

3.4.1 Cesión de datos

La cesión de datos implica compartir información con otras entidades para que sea utilizada, almacenada, modificada o transformada según la finalidad específica establecida. Este proceso también conlleva una **transferencia de responsabilidad** en cuanto al uso adecuado de los datos, lo que implica que la entidad receptora asume la responsabilidad de cumplir con los fines acordados y las normativas vigentes en protección de datos.

Asimismo, cuando la transferencia de datos consiste en una cesión, el responsable del tratamiento a quien se ceden los datos personales queda sujeto a las **mismas obligaciones legales y reglamentarias** que el responsable cedente.

La cesión puede ser para pedidos únicos, recurrentes o periódicos, y la modalidad de transferencia varía en función del uso que se le quiera dar.

- **Cesión para explotación y análisis de datos:** los datos se ceden para realizar análisis de calidad, crear tableros de explotación o desarrollar modelos predictivos. Este tipo de cesión permite modelar y almacenar los datos en las bases del receptor, siempre cumpliendo con los principios de protección de datos.

- **Cesión para almacenamiento y resguardo de datos:** el almacenamiento y resguardo de datos tiene como objetivo preservar la integridad de los datos a largo plazo, sin posibilidad de modificación. Esta finalidad es común en situaciones donde los datos deben ser archivados por razones legales, históricas o cumplimiento normativo, tales como auditorías tanto internas como externas.

En el GCABA, la cesión de datos que no requiere consentimiento se realiza bajo acuerdos claros que definen los límites de uso, asegurando que no se desvíe de los fines acordados ni perjudique a los titulares de la información. Para formalizar este proceso, las organizaciones que requieran la cesión de datos podrán utilizar el *Modelo de Requerimiento para la Cesión de Datos* según la Resolución N°136-SECITD/22.

En los casos donde se requiera consentimiento, también se deberá presentar el consentimiento del titular de los datos, al que se le debe informar sobre la finalidad de la cesión.

🔍 **Modos de envío para cesión de datos**

Existen diversas formas de enviar los datos, dependiendo de las necesidades y del tipo de cesión. Los datos pueden ser cedidos mediante accesos controlados a repositorios o bases de datos, o transmitidos en intervalos regulares para alimentar tableros de control, lo que permite el monitoreo de indicadores en tiempo real.

Es fundamental garantizar la seguridad y la confidencialidad de la información, para ello se debe utilizar exclusivamente canales y plataformas oficiales, evitando el uso de herramientas personales. Además, los datos deben enviarse únicamente a destinatarios específicos, limitando el acceso a las personas estrictamente necesarias para cada caso.

En casos donde se cuente con datos personales o sensibles, es importante también encriptar la información, resguardarla con contraseña y asegurarse de que se entregue en los formatos y estructuras previamente definidos, manteniendo así su integridad y utilidad.

3.4.2. Interoperabilidad de sistemas

La interoperabilidad permite realizar consultas a demanda y en tiempo real a través de sistemas automatizados, mediante servicios web. Este tipo de transferencia garantiza que los datos se obtengan directamente del sistema que los provee, lo que asegura que siempre estén actualizados.

En el GCABA, la interoperabilidad de datos debe realizarse a través del **Sistema de Interoperabilidad del Gobierno de la Ciudad de Buenos Aires**, denominado X-BA, creado bajo el Decreto N° 118/22 y reglamentado por la Resolución N° 303/22 y su modificatoria N° 236/23.

Cuando se trata de la publicación de un servicio web para disponibilizar **datos en formato abierto**, no es necesario utilizar X-BA, ya que no requiere la capa de seguridad y gobernanza de datos que este sistema aporta. Sin embargo, para el **intercambio de datos públicos, personales o sensibles, se recomienda utilizar X-BA,** aunque con una operatoria diferenciada según el tipo de dato y las condiciones requeridas para su transferencia.

Es necesario que tanto el emisor como el consumidor de los datos sean **miembros del sistema de interoperabilidad X-BA** para poder llevar a cabo la transferencia. Para habilitar el consumo del dato, el receptor debe utilizar el **Portal de Gestión de Servicios de Interoperabilidad**, donde se le solicitará la justificación de competencias o interés legítimo, el nivel de consumo y la integración específica, es decir, el alcance de utilización de ese servicio.

Asimismo, si para compartir el dato es necesario solicitar el consentimiento de su titular, según las consideraciones de las secciones anteriores, el emisor del dato es responsable de generar una cadena de confianza que permita obtener dicho consentimiento informado. Para las organizaciones miembro que utilizan X-BA, este consentimiento podrá ser solicitado al titular de los datos a través de la plataforma miBA. Una vez obtenido el consentimiento, se activará automáticamente la respuesta.

🔍 Modos de envío para interoperabilidad

Para la interoperabilidad de datos, se requiere el uso de servicios web como canal de transferencia. La habilitación para acceder al servicio web y realizar consultas se concede una única vez mediante el Portal de Gestión de Servicios de Interoperabilidad. Una vez otorgado el acceso, la organización podrá consultar los datos necesarios para el trámite o procedimiento específico.

Es fundamental que la información obtenida no sea almacenada en las bases de datos del receptor. Cada vez que se necesite acceder a la información, se deberá realizar una nueva consulta al sistema, garantizando que los datos sean siempre actualizados y precisos. A diferencia de la cesión de datos, en la que los datos pueden

ser almacenados y modificados, en la interoperabilidad el uso es limitado, **no está permitido guardar ni modificar la información consultada**, ni podrá reutilizarse para la validación de un dato en un proceso posterior. Asimismo, los datos obtenidos por la interoperabilidad no podrán cederse a terceros.

No obstante, es importante que el resultado de las consultas sea registrado en un documento específico con el fin de facilitar auditorías o revisiones posteriores. Esto asegura la transparencia y trazabilidad del proceso, manteniendo al mismo tiempo la integridad y actualidad de la información consultada.

3.4.3. Pedidos y publicación de datos

La **transferencia de datos con fines de publicación** o por solicitud puntual permite que la información pública esté disponible para su reutilización por parte de ciudadanos, organizaciones y otras entidades. Estos mecanismos son esenciales para fomentar la transparencia, el acceso a la información y el aprovechamiento de los datos para fines diversos, desde investigaciones y proyectos hasta el desarrollo de soluciones innovadoras.

El GCABA facilita el acceso a datos a través de procesos que garantizan el cumplimiento de normativas y la protección adecuada, dependiendo del tipo de solicitud y la clasificación de los datos. Existen diferentes enfoques para acceder a esta información, que varían en función de la frecuencia de acceso y el uso permitido, asegurando que los datos estén disponibles de manera segura y accesible para quienes los requieran.

3.4.3.1. Pedidos puntuales de datos

En el **GCABA**, los pedidos puntuales de datos están regulados por la **Ley N° 104**, que garantiza el derecho de acceso a la información pública. Esto permite a los ciudadanos solicitar cualquier dato creado u obtenido por el órgano requerido, y que se encuentre en su posesión y bajo su control, con las excepciones señaladas en las secciones anteriores.

Este tipo de transferencia permite que ciudadanos u organizaciones soliciten acceso a datos específicos a través de un proceso único, para el cual no es necesario acreditar derecho subjetivo ni interés legítimo, solo se deben

cumplir con las formalidades del artículo N° 9 de la Ley N° 104 de Acceso a la Información Pública, que deben incluir en la solicitud:

- Proveer nombre, apellido y datos de contacto.
- Constituir domicilio en la Ciudad Autónoma de Buenos Aires o proporcionar un correo electrónico válido.
- Especificar claramente la información solicitada.

En el caso de que la información requiera consentimiento, será necesario anonimizar los datos antes de su entrega o presentar el consentimiento del titular.

La solicitud se realiza una única vez, y si se requiere la misma información en el futuro, es necesario presentar una nueva solicitud. Los sujetos obligados por la Ley N° 104 de Acceso a la Información Pública, y en el marco del Plan de Transparencia Activa, deberán publicar en sus respectivas páginas web, de manera completa y actualizada y en lo posible en formatos abiertos y reutilizables, la información que se exige por normativa.

Las solicitudes de acceso a la información y sus respuestas, incluyendo la información entregada, serán públicas, debiendo el GCABA poner a disposición del público esta información una vez contestado el pedido de acceso a la información.

Como buena práctica, cuando la información solicitada es recurrente, se recomienda su publicación en formato abierto en el **Portal de Datos Abiertos de la Ciudad, Buenos Aires Data,** para facilitar su acceso continuo y estableciendo su frecuencia de actualización.

3.4.3.2. Publicación en formato abierto

La **publicación de datos en formato abierto** permite el acceso continuo y público a la información, promoviendo la transparencia y facilitando la reutilización por parte de cualquier usuario, incluidos ciudadanos, organizaciones de la sociedad civil, empresas y otros sectores interesados. Estos datos, disponibles en el **Portal de Datos Abiertos de la Ciudad, Buenos Aires Data**, pueden ser utilizados para diversos fines, como proyectos personales, comerciales, investigaciones académicas o el desarrollo de aplicaciones tecnológicas.

El **Portal Buenos Aires Data** es la plataforma oficial implementada por el Gobierno de la Ciudad Autónoma de Buenos Aires bajo el **Decreto N° 156/2012** para facilitar la búsqueda, descubrimiento y acceso a conjuntos de datos del sector público.

La publicación de estos datos es permanente, lo que permite su consulta en cualquier momento. La entidad emisora del dato, es decir, la fuente auténtica, define la frecuencia y el tipo de datos que se publican, mientras que la Dirección General de Gobernanza de Datos establece el formato de publicación.

Con el objetivo de estandarizar la información y facilitar su explotación, se utiliza el **Diccionario de Datos y Metadatos del Gobierno de la Ciudad**, que define el formato y las limitaciones de cada dato. Cada conjunto de datos publicado incluye un catálogo detallado con los campos del recurso, su formato y una descripción clara, lo que permite a los usuarios comprender y reutilizar la información de manera eficiente. Para profundizar sobre el Diccionario de Datos y Metadatos, se puede consultar el Anexo I – Diccionario de Datos y Metadatos, de estos lineamientos.

Es importante aclarar que en este formato se publican únicamente **datos públicos o datos personales anonimizados**. Por ejemplo, se publican conjuntos de datos como el número de nacimientos, fallecimientos, o viajes en subte y bicicleta, donde cualquier dato identificable ha sido anonimizado previamente para garantizar la protección de la privacidad.

🔍 Modos de envío para su publicación

Para asegurar la frecuencia establecida en la publicación de los datos y garantizar que la estructura de los datos se mantenga consistente, se recomienda adoptar un modelo automatizado mediante un proceso de Extracción, Transformación y Carga (ETL) entre la base de datos y el **Portal de Datos Abiertos**.

En caso de no contar con una fuente de datos sistematizada, se permitirá el envío de archivos a través de un servidor FTP proporcionado por el operador de Buenos Aires Data. Para mantener la cadena de confianza, se desalienta el uso de archivos almacenados en nubes personales o el envío de información a través de correos electrónicos. Se sugiere iniciar un plan para sistematizar la información, lo que facilitará la implementación de procesos automáticos en el futuro.

Este enfoque contribuye a garantizar que se cumplan las frecuencias establecidas para la publicación de cada conjunto de datos y que la estructura de datos sea siempre la misma.

3.5. Ejemplos

3.5.1. Pedidos en el marco de la Ley N° 104

Una empresa periodística envía un pedido de información en el marco de la Ley N° 104 de Acceso a la Información Pública, solicitando nombre y apellido, enfermedades laborales y adhesión sindical de las personas funcionarias de su área.

¿Qué hay que hacer? En este caso, se deberá dar respuesta al pedido de información por obligación legal, pero **no se deberán detallar todos los datos solicitados**. Los únicos datos que es posible informar serán **nombres y apellidos** ya que, al figurar en registros públicos estatales por pertenecer a la dotación de la repartición consultada, estos datos son considerados públicos.

Los datos correspondientes a la salud y afiliación a sindicato no pueden transferirse por ser considerados **datos sensibles**. Esos datos sí podrían ser compartidos si, por ejemplo, **se solicita el porcentaje total de personas funcionarias afiliadas a sindicatos, o algún dato anónimo.** En esos casos, al no revelarse la identidad de las personas, no habría inconveniente.

3.5.2. Pedidos de información entre áreas gubernamentales

El Ministerio de Educación solicita al Ministerio de Salud validar los datos de la “Libreta de Salud Escolar” o en su defecto el “Certificado de vacunas” de un/una estudiante para su inscripción escolar. En este sentido, dentro de las competencias otorgadas mediante el Decreto N° 387/23 y sus modificatorios y en el marco del Programa de Salud Escolar, intercambiar esta información se encontraría dentro de las misiones y funciones de las reparticiones.

Como vimos en las clasificaciones, si bien se trata de datos personales y relativos a la salud, el artículo N° 7 de la Ley N° 1.845 de Protección de Datos Personales establece, dentro de las excepciones al requisito del consentimiento para el tratamiento de los datos personales, que los mismos **se recaben para el ejercicio de funciones propias de los poderes del Gobierno** de la Ciudad de Buenos Aires, o en virtud de una obligación legal.

Es por ello que los datos podrán ser compartidos **sin consentimiento previo** del titular, en virtud de los parámetros allí establecidos.

Ahora bien, si el objetivo es realizar un **cruce y análisis de datos** para desarrollar estrategias de salud escolar, la mejor opción sería la **cesión de datos**. Esto permite obtener un conjunto de datos completo que puede ser almacenado y modelado en las bases del receptor para llevar a cabo análisis profundos y poder tomar decisiones basadas en evidencia.

Por otro lado, si se busca establecer un procedimiento para **verificar el dato al momento de hacer la tramitación,** la mejor alternativa es la **interoperabilidad**. Este método permite **validar la información en tiempo real** a través de consultas puntuales.

4. CALIDAD DE DATOS

4.1. Definición e importancia de la calidad de datos

Llamamos **lineamientos de calidad de datos** a un conjunto de *normas y procedimientos* que establece una organización para asegurar que los datos que utiliza cumplan con ciertos *criterios de calidad*, garantizando que los datos sean precisos, consistentes, completos, oportunos y relevantes para los propósitos de la organización.

Establecer y seguir **lineamientos de calidad de datos** es fundamental para alcanzar y mantener altos estándares de calidad en los datos. Estos lineamientos proporcionan un marco estructurado que ayuda a prevenir y corregir problemas de calidad, asegurando que los datos se gestionen de manera coherente y eficiente en toda la organización. Además, facilitan la auditoría y el cumplimiento normativo, minimizan el riesgo a cometer errores, y garantizan que todos los usuarios de datos dentro de la organización trabajen con un entendimiento común y alineado sobre lo que constituye datos de calidad.

La **calidad de datos** se refiere al grado en el que los datos y la metadata cumplen con las necesidades de los usuarios y se ajustan a ciertos *criterios o dimensiones* que veremos en detalle más adelante. De momento, debemos saber que los datos calificados como de '*alta calidad*' son útiles para describir la situación actual en torno a una problemática, analizar tendencias, tomar decisiones y cumplir objetivos estratégicos.

Así como la calidad de los datos es un factor clave en cualquier organización, la determinación de la relevancia y utilidad de los datos dependerá del contexto en el que se utilicen. De esta manera, pueden existir datos que sean significativos para un organismo de gobierno y, por tanto, sea imprescindible que alcancen determinados niveles de calidad, pero quizás no lo sean para otra dependencia.

Contar con datos de calidad es esencial para cualquier organización que desee tomar decisiones informadas y basadas en evidencia. Los datos de baja calidad pueden llevar a conclusiones erróneas y toma de decisiones ineficaces, lo que puede traducirse en costos financieros y deterioro de la confianza en la entidad. En cambio, los datos de alta calidad permiten a las organizaciones optimizar sus operaciones, identificar oportunidades de mejora y mitigar riesgos.

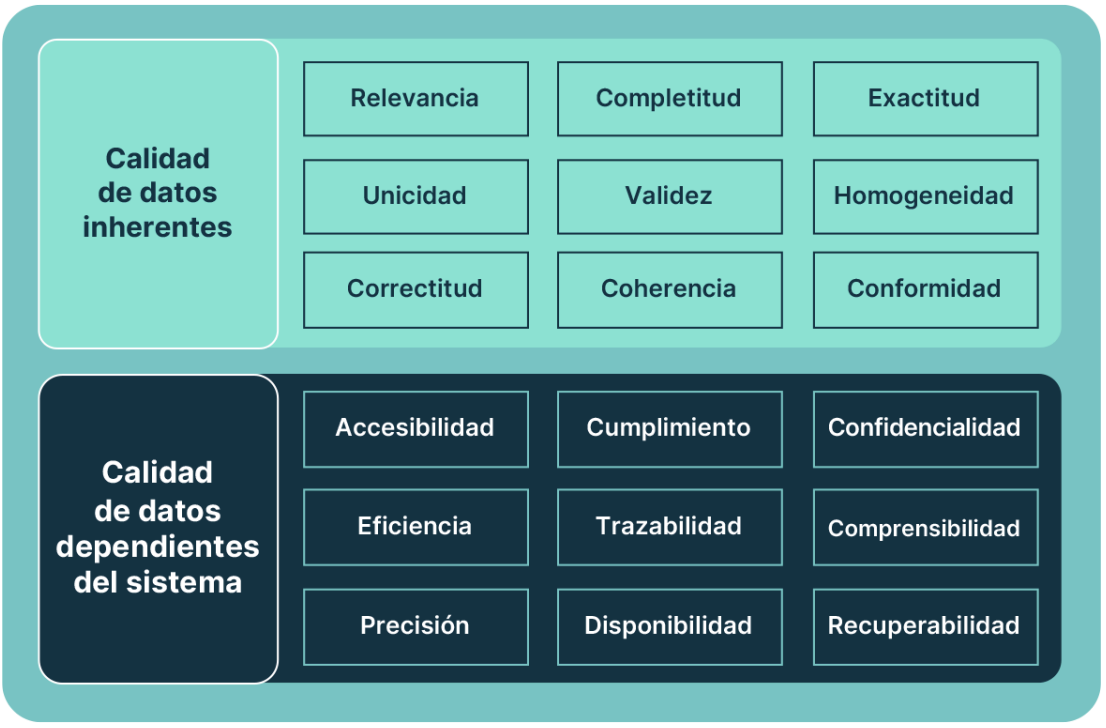
4.2. Criterios de calidad

La **Organización Internacional de Normalización (ISO)** y la **Comisión Electrotécnica Internacional (IEC)** definieron en el documento **ISO/IEC 25012** un modelo general de calidad para los datos representados en un formato estructurado en un sistema informático. A su vez, se incluyeron criterios propios en la adopción de esta estructura de análisis para incluir ejes orientados a la gestión de datos del Gobierno de la Ciudad Autónoma de Buenos Aires.

La calidad de los datos puede clasificarse en los siguientes grupos:

1. **Calidad de datos inherentes**: es el grado con el que las características de calidad de los datos tienen el potencial intrínseco para satisfacer las necesidades establecidas. Abarca los valores de dominios de datos y posibles restricciones, reglas de modelo de negocio, las relaciones entre valores de datos y los metadatos.

2. **Calidad de datos dependientes del sistema**: refiere al grado con el que se alcanza y preserva la calidad de datos mediante un sistema informático cuando los datos se usan bajo condiciones específicas. Desde este punto de vista, la calidad de datos depende de las capacidades de los componentes del sistema informático.



En esta guía, nos basaremos en la **calidad de datos inherentes** para la elaboración de dimensiones de análisis. La correspondiente clasificación de los problemas de datos en alguna de estas dimensiones, facilitarán el diagnóstico y el tratamiento a aplicar a las fuentes de origen para una gestión más eficiente de los datos.

4.3. Análisis

4.3.1. Análisis preliminar de la estructura de datos

4.3.1.1. Fuente de origen de los datos

La fuente de origen de los datos es de dónde proviene la información: pueden ser datos públicos consumibles desde una página web, archivos en formato CSV, Excel, JSON, PDF, Google Sheets, bases de datos relacionales (Oracle Database, MySQL, SQL Server, SQLite, DBeaver, etc), bases de datos no relacionales (NoSQL), API's, entre otros. Cuál sea la fuente de información y el formato de los datos son aspectos que van a condicionar cómo aplicar las dimensiones de calidad que veremos más adelante.

En esta guía nos centraremos en el análisis de calidad de **datos estructurados**. Los datos estructurados son aquellos organizados en un formato fijo y predefinido, como tablas, donde la información se almacena en filas y columnas con tipos de datos específicos. Esto facilita su almacenamiento, consulta y análisis. Archivos en Excel, hojas de cálculo, o SQL son ejemplos de ello.

Es importante recordar que la información contenida en archivos de tipo Excel u hojas de cálculo se organiza en tablas con filas (registros) y columnas (campos). En las bases de datos relacionales existen entidades (cada una constituye un objeto o concepto) con sus atributos (propiedades de las entidades) y relaciones (forma de asociación entre entidades). Cada entrada de datos constituye un registro.

Una vez relevado e identificado de dónde proviene la información y cuál es su estructura, pasaremos al análisis de la disponibilidad de recursos.

4.3.1.2. Disponibilidad de recursos

Antes de iniciar un análisis de calidad de datos, es fundamental evaluar la información y los recursos disponibles en torno a los datos y la estructura de datos que se pretenden peritar.

Este paso es imprescindible dado que permitirá la correcta interpretación de los datos y, como tal, condiciona todo el análisis subsiguiente que se realice sobre los mismos. Realizar esta evaluación previa garantiza que el análisis de calidad se lleve a cabo con una comprensión completa y precisa de los datos, reduciendo el riesgo de malas interpretaciones y errores durante el proceso.

Esta evaluación de la disponibilidad de recursos incluye tres aspectos clave:

4.3.1.2.1. Conocimiento del DER (Diagrama Entidad-Relación)

Cuando nos enfrentamos al análisis de una base de datos relacional, es esencial comprender su estructura, incluyendo las claves primarias (PK) y claves foráneas (FK), así como la cardinalidad de las tablas. Este conocimiento asegura que la base de datos sea estructuralmente coherente, que las relaciones entre los diferentes elementos de datos sean claras y posibilita identificar inconsistencias o errores en la integración de datos.

Este análisis permite identificar, por ejemplo, si la relación entre entidades es o no óptima. La cardinalidad entre tablas será óptima cuando sus elementos se relacionen de uno a uno (1:1) o de uno a muchos (1:N) o de muchos a uno (N:1), pero nunca si la relación es de muchos a muchos (N:N), dado que ese caso exige la creación de tablas intermedias para desambiguar dichas relaciones.

4.3.1.2.2. Disponibilidad de una guía de datos

Toda fuente de origen que contenga entidades y atributos será susceptible de contar con una guía de datos asociados. Contar con este recurso es un aspecto de suma importancia para comprender el significado de los datos y metadatos, su formato, restricciones, y su propósito dentro del sistema. Esta información habilita que los datos se utilicen de manera adecuada según su definición.

Una guía de datos es un documento o repositorio que describe los elementos de una base de datos. Esto incluye la descripción acerca de qué

datos están almacenados, cuál es su formato, su longitud, cuáles son los valores permitidos o las restricciones que pesan sobre ellos, de dónde provienen, cuál es su frecuencia de actualización, entre otros aspectos.
Los metadatos consisten en información que caracteriza datos, describen el contenido, calidad, condiciones, historia, disponibilidad y otras características.

La guía de datos de cada fuente de origen debe encontrar su fundamento en los estándares publicados en el **Diccionario de Datos y Metadatos del Gobierno de la Ciudad Autónoma de Buenos Aires,** según lo detallado en el Anexo I de esta guía.

4.3.1.2.3. Documentación Funcional del Modelo de Negocios

Es necesario tener acceso a la documentación funcional que describe el modelo de negocios en detalle. Esta documentación proporciona contexto sobre cómo los datos se relacionan con los procesos y objetivos de la organización, permitiendo una interpretación alineada con las necesidades y expectativas del negocio.

El modelo de negocios constituye el marco conceptual que describe cómo una organización crea, entrega y captura valor a través de sus operaciones y procesos. Es una representación detallada de los componentes clave de la organización, como sus productos o servicios, clientes, recursos, flujos de ingresos y costos, y cómo estos interactúan entre sí. Comprender el modelo de negocios es lo que permite entender cuál es el valor que aportan los datos y qué se apunta a conseguir con ellos.

4.3.2. Análisis de calidad de datos

4.3.2.1. Evaluación de dimensiones

Las dimensiones de calidad de datos constituyen un eje fundamental en el diagnóstico de problemas que podemos acarrear en una fuente de origen determinada. Se pueden aplicar infinidad de dimensiones en función de las reglas específicas del modelo de negocio de la fuente de datos. Sin embargo, en esta guía se fundamentan 9 dimensiones de calidad de datos inherentes que satisfacen criterios propios de las fuentes de datos del Gobierno de la Ciudad Autónoma de Buenos Aires.

4.3.2.1.1.Relevancia

Ya completado el análisis preliminar de las fuentes de origen de los datos y los recursos disponibles en torno a ellos, nos centraremos en la primera dimensión de calidad de datos: la **relevancia**.

En este contexto, los datos relevantes son los que realmente importan según las necesidades específicas de los usuarios que los utilizan. En otras palabras, los datos y la estructura de datos son considerados relevantes cuando son útiles y significativos para el análisis, la toma de decisiones o el cumplimiento de objetivos.

🔍 **Fuentes de origen de datos**

La relevancia aplica a todas las fuentes de datos.

💡 **Resultados esperados**

Como resultado de este acercamiento al conjunto de datos, debemos ser capaces de individualizar y dejar asentado cuáles son las entidades y atributos relevantes para el análisis y por ende priorizables que establezcan la hoja de ruta para aplicar las subsiguientes dimensiones de calidad.

$$\frac{Atributos\ Relevantes}{Atributos\ Totales} = Puntaje\ de\ relevancia$$

✅ **Ejemplos**

La siguiente tabla contiene información sobre **Postas Digitales.** A partir del análisis de la documentación funcional del modelo de negocio, se definió que los atributos útiles para los objetivos son el ID, el nombre de la posta, descripción del servicio, fecha de la última verificación y estado. Las columnas de observaciones y código postal donde se encuentra instalado el servicio no cumplen con los criterios de relevancia, por lo que no serán evaluados en las siguientes etapas de calidad.

Relevancia

ID	nombre_posta_d	codigo_postal	descripcion_servicio	fecha_ultima_verificacion	estado	observaciones
1	Posta Digital Belgrano	A1234	Wi-fi gratuito y acceso a computadoras públicas	2023-08-14	Funcionando	No se presentan irregularidades
2	Posta Digital Palermo	B5678	Asistencia en trámites digitales	2023-08-15	No funcionando	Vandalizado, con averiguación a cargo de las autoridades
3	Posta Digital Almagro		Capacitación en herramientas digitales	2023-08-16	En reparación	A la espera de la importacion de materiales
4	Posta Digital Caballito	U1122	Impresión 3D y asesoría en tecnología	2023-08-17	Parcialmente funcionando	No se presentan daños, insumos limitados
5	Posta Digital Villa Urquiza	A1345	Aula virutal para educación a distancia	2023-08-18	Funcionando	No se presentan irregularidades
6	Posta Digital San Telmo	A1234	Asesoría en comercio electrónico	2023-08-19	No funcionando	
7	Posta Digital Flores		Registro de identidad digital	2023-08-20	Funcionando	No se presentan irregularidades
8	Posta Digital Retiro	C9101	Centro de innovación y desarrollo tecnológico	2023-08-21	En reparación	
9	Posta Digital Barracas	U1122	Capacitación en programación y desarrollo web	2023-08-22	Parcialmente funcionando	
10	Posta Digital Recoleta	A1345	Biblioteca digital y acceso a e-books	2023-08-23	Funcionando	Queda en suspenso hasta proxima visita

Análisis

Atributo	Relevancia
id	*Si*
nombre_posta_d	*Si*
codigo_postal	*No*
descripcion_servicio	*Si*
fecha_ultima_verificacion	*Si*
estado	*Si*
observaciones	*No*

4.3.2.1.2. Completitud

La **completitud** evalúa que los datos no sean nulos en un atributo que requiera estar completo según sus reglas de negocio, es decir, que no falten datos críticos.

🔍 **Fuentes de origen de datos**

La completitud aplica a todas las fuentes de origen de datos, pero sólo sobre *aquellos atributos que en el punto anterior (Relevancia) se identificó que requieren estar completos*.

💡 **Resultados esperados**

El output esperado es un porcentaje que refleje el nivel de completitud de los atributos que tengan la condición de obligatoriedad de ser completos o no nulos.

$$\%\ Completitud\ Atributo = \frac{Datos\ completos}{Datos\ totales}$$

$$n = total\ de\ atributos\ que\ calif ican\ para\ evaluar\ completitud$$

$$\%\ Completitud\ Total = \frac{Compl\ Atrib_1 + Compl\ Atrib_2 + Compl\ Atrib_3 + \ldots + Compl\ Atrib_n}{n}$$

✅ Ejemplos

Los Identificadores y claves primarias (PK) son atributos que siempre exigen estar completos. La completitud de otros campos dependerá de las reglas de negocio. Como ejemplo, suponiendo que tenemos una tabla cuyo fin es individualizar tarjetas de crédito y la cuenta bancaria a la que se encuentra asociada, es de esperar que el número del plástico sea un atributo que no acepte registros nulos.

En la tabla a continuación es posible encontrar un listado de distintos servicios que se brindan desde el Gobierno de la Ciudad de Buenos Aires. En color se encuentran resaltados los registros que no cumplen con el criterio de completitud establecido.

	ID	Servicio	Descripción	Área Gubernamental
✓	13482	HealthAI	Plataforma de diagnóstico médico preventivo mediante análisis de big data.	Salud Pública
✓	27891	EduSmart	Sistema de aprendizaje personalizado para estudiantes de escuelas públicas.	Educación
	34675	SafeCity	Monitoreo de seguridad urbana con predicción de incidentes en tiempo real.	Seguridad Ciudadana
✗		EcoAI	Optimización del consumo energético en edificios públicos mediante AI.	Medio Ambiente y Energía
✓	51932	TaxOptimizer	Servicio automatizado para la optimización de la recaudación de impuestos.	Finanzas
✓	63784	TransAI	Gestión inteligente de tráfico urbano y transporte público.	Transporte y Tránsito
	75413	✗	Asistencia jurídica automatizada para consultas legales básicas.	Justicia y Derechos Humanos
✓	86125	CitizenConnect	Plataforma de atención ciudadana con chatbots para consultas y reclamos.	Atención al Ciudadano
	90457	AgroVision	Monitoreo y predicción de cosechas para agricultores mediante imágenes satelitales.	✗
✓	10234	WorkMatch	Plataforma de empleo público con recomendación automática de vacantes.	Empleo y Recursos Humanos

Vacíos o nulos en **Claves Primarias, o identificadores**

Nulos o vacíos en categorías de atributos necesarios según las **definiciones de las reglas de negocio**

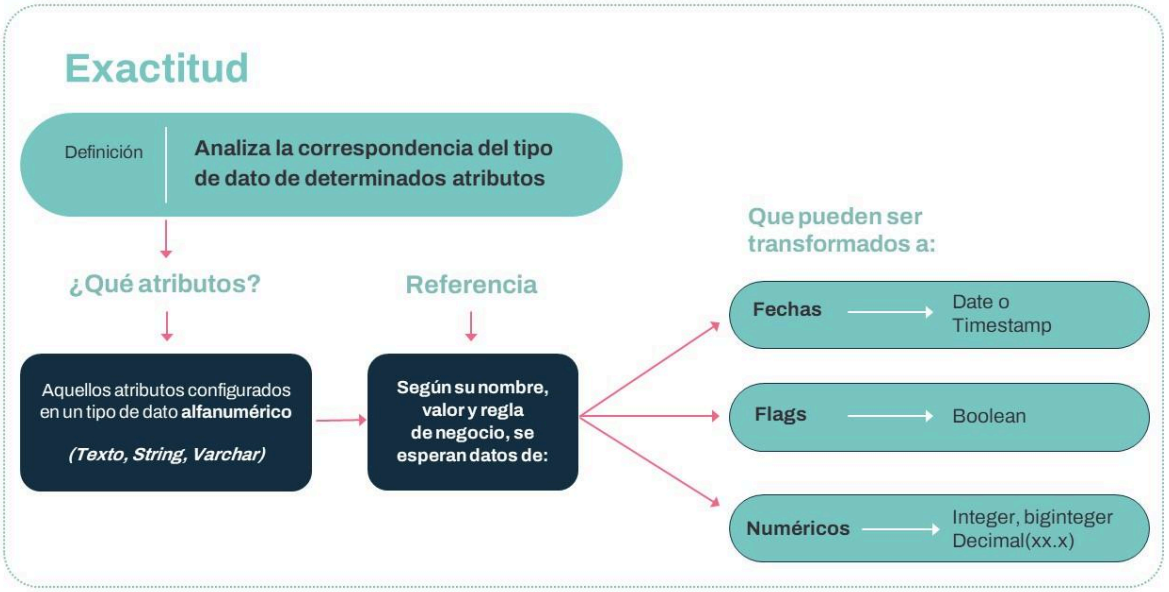
4.3.2.1.3.Exactitud

La **exactitud** consiste en la adecuación del tipo de dato que se espera para un atributo específico. El análisis de la exactitud debe ocurrir posterior al de completitud e implica dos acciones:

a. Corroborar que los atributos coincidan con el tipo de dato esperado. Por ejemplo, si existe un campo “fecha_alta” que tenga todos sus registros en formato fecha; si existe un campo llamado “precio” que tenga todos sus registros en formato numérico; si existe un campo dicotómico para indicar una relación verdadero/falso, SI/NO, etc; que todos sus registros sean de tipo *booleano*, si existe un campo “nombre”, que tenga todos sus registros en formato texto/string/varchar.

> **En bases de datos estructuradas, cada atributo debe estar definido desde el origen en un determinado formato. Por tal motivo, lo esperable sería no tener inconvenientes con respecto a este ítem. En cambio, cuando la fuente de origen es un texto plano (txt, csv, tsv, json), sí corresponde evaluar cada atributo en detalle y asignarle el formato correcto.**

b. Para los atributos que debiendo ser fechas, numéricos o booleanos se encuentren en formato alfanumérico, deben poder convertirse al formato que corresponde según el modelo de negocio.



🔍 **Fuentes de origen de datos**

Para archivos en **texto plano**, se debe evaluar la exactitud de todos los campos. Para **datos estructurados**, se debe evaluar los atributos en formato alfanumérico (texto/string/varchar) que debieran estar en formato fecha, numérico o booleano.

💡 **Resultados esperados**

Si bien es necesario comenzar con el punto **(a)** para conocer en cuáles atributos se debe analizar la factibilidad de su conversión, el output esperado es un porcentaje que refleje el nivel de exactitud del punto **(b)**.

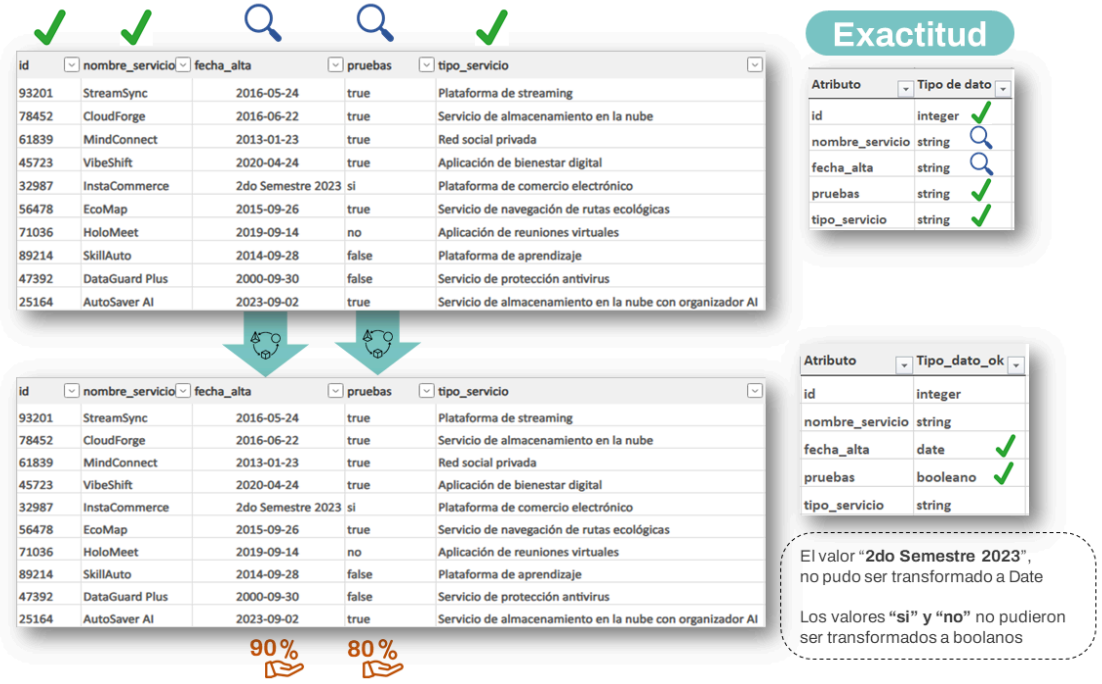
$$\%\ Exactitud\ Atributo = \frac{Datos\ exactos}{Datos\ totales}$$

$$n = total\ de\ atributos\ que\ calif\ican\ para\ evaluar\ exactitud$$

$$\%\ Exactitud\ Total = \frac{Exact\ Atrib_1 + Exact\ Atrib_2 + Exact\ Atrib_3 + \ldots + Exact\ Atrib_n}{n}$$

✅ **Ejemplos**

En el siguiente esquema de una base relacional en SQL observamos una tabla con sus atributos y qué tipos de datos asumen. Comprobamos que algunos atributos son de tipo string, si bien la lógica indica que debieran ser de otro formato. Mediante las conversiones pertinentes, se verifica que dichos datos pueden convertirse al tipo que corresponde.



4.3.2.1.4. Unicidad

El análisis de **unicidad** evalúa la presencia de duplicados y la comprobación de valores únicos según la lógica del modelo de negocio. Implica dos acciones:

a. Corroborar que todas las claves primarias (PK) o identificadores sean únicos.
b. Corroborar que los atributos que según las reglas de modelo de negocio o indicación de la guía de datos deban ser únicos efectivamente lo sean.

🔍 Fuentes de origen de datos

En las bases de datos relacionales, los identificadores o claves primarias (PK) son únicos, por lo que el punto **(a)** estaría solucionado y sólo habría que corroborar el cumplimiento de los datos con respecto al punto **(b)**. Esto no ocurre en los archivos de texto plano, por lo que hay que prestar especial atención a la corroboración de la unicidad en ese tipo de fuente de datos.

💡 Resultados esperados

Se debe obtener un porcentaje de ambos puntos, **(a)** y **(b)**.

$$\%\ Unicidad\ Atributo = \frac{Datos\ \acute{u}nicos}{Datos\ totales}$$

$$n = total\ de\ atributos\ que\ calif ican\ para\ evaluar\ unicidad$$

$$\%\ Unicidad\ Total = \frac{Unic\ Atrib_1 + Unic\ Atrib_2 + Unic\ Atrib_3 + \ldots + Unic\ Atrib_n}{n}$$

✅ Ejemplos

Imaginemos una entidad que sea **paises_del_mundo**. Las reglas de negocio indican que, en la tabla correspondiente, los países no pueden estar repetidos, por lo cual cualquier duplicado significaría un error en la unicidad de los registros. Ahora supongamos que tenemos otra tabla en la que se registran las ciudades más pobladas del mundo, aquellas con más de un millón de habitantes. Para este caso, es posible encontrar países repetidos, ya que un país puede acunar varias ciudades de más de un millón de habitantes. En este caso, el modelo de negocio nos determina que el atributo "*país*" puede aparecer repetido, pero no así el atributo "*ciudad*".

Para el ejemplo a continuación, se grafica una tabla llamada "barrios", en la cual los barrios de la Ciudad Autónoma de Buenos Aires no pueden estar repetidos. Por lo tanto, se resalta en color todas las veces que tales atributos o los identificadores no cumplen con el criterio de unicidad. Se observa que la Comuna correspondiente a cada barrio, así como la Ciudad y el país, sí pueden aparecer repetidos de acuerdo con la lógica del negocio.

4.3.2.1.5. Validez

Al analizar la **validez** estamos evaluando la autenticidad y confiabilidad de los registros de una entidad en base a una **fuente de validación externa**.

Llamamos fuente de validación externa a una base o registro elaborada y mantenida por otra dependencia que contenga un conjunto de datos que sea comparable a los que estamos evaluando.

Un ejemplo de fuente validadora es el Registro Nacional de las Personas (RENAPER) o el Sistema de Identificación Nacional Tributario y Social (SINTyS), el cual contiene información patrimonial y social de las personas físicas y jurídicas. Otro ejemplo podría ser el Registro Nacional de Sociedades, disponible online para consultar públicamente.

De tal manera, evaluar la validez significará corroborar que los atributos que cuenten con una fuente de validación externa sean auténticos.

🔍 **Fuentes de origen de datos**

La validez aplica a toda fuente de origen de los datos, siempre y cuando sea posible validar el atributo con una fuente validadora. Conviene tener presente que no todos los atributos admitirán posibilidad de validación externa.

💡 Resultados esperados

Se debe obtener un porcentaje, siendo el objetivo esperado lograr el **100 %**. Cualquier número inferior a esa cifra debería considerarse una alerta.

$$Dato\ vá\ lido = Dato \in Fuente\ validadora$$

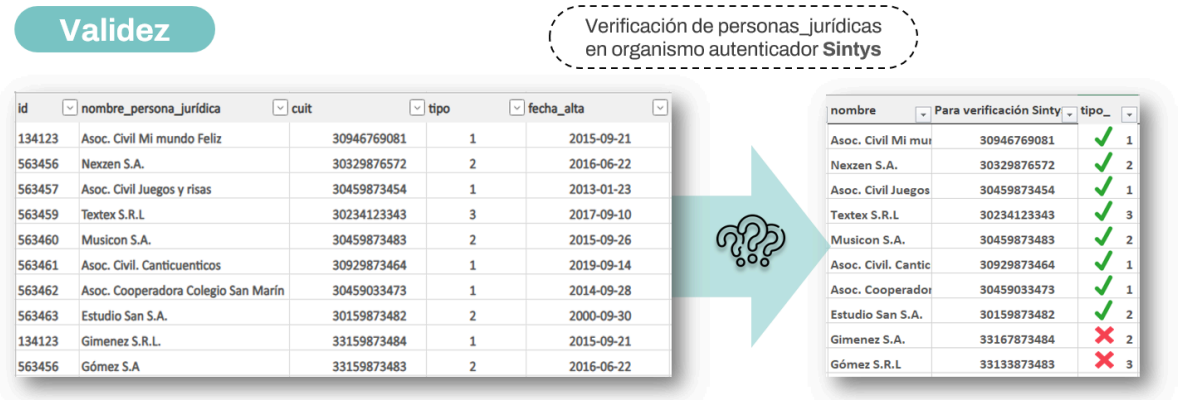
$$\%\ Validez\ Atributo = \frac{Datos\ vá\ lidos}{Datos\ totales}$$

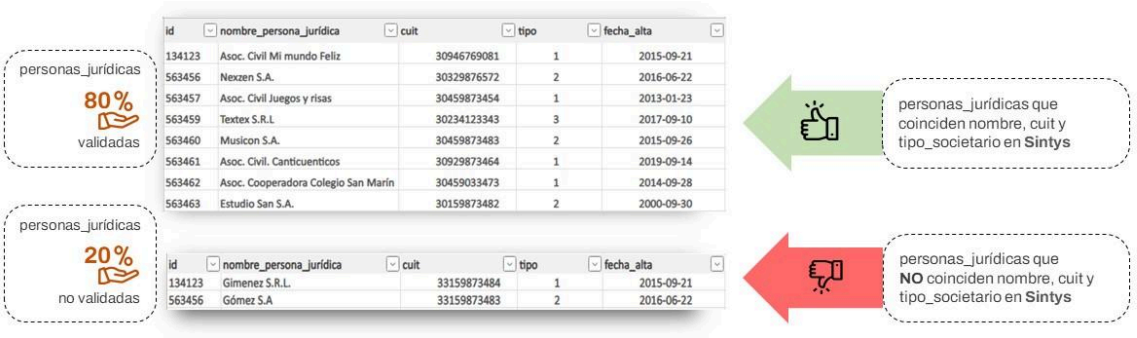
$$n = total\ de\ atributos\ que\ calif\ican\ para\ evaluar\ validez$$

$$\%\ Validez\ Total = \frac{Valid\ Atrib_1 + Valid\ Atrib_2 + Valid\ Atrib_3 + \ldots + Valid\ Atrib_n}{n}$$

✅ Ejemplos

Para la validación de las personas físicas en este ejemplo vamos a utilizar la base del **Sistema de Identificación Nacional Tributario y Social** (**SINTyS**), cruzando con el campo num_documento.





4.3.2.1.6. Homogeneidad

La dimensión de **homogeneidad** evalúa si los datos están uniformemente clasificados, de acuerdo a las categorías predefinidas según la guía de datos.

🔍 **Fuentes de origen de datos**

La homogeneidad aplica a toda fuente de origen de datos, pero solamente a los atributos que sean *categorizables*.

💡 **Resultados esperados**

Se debe obtener un porcentaje de los atributos estandarizables, siendo el objetivo esperado lograr el **100 %**. Cuanto mayor sea el resultado obtenido en homogeneidad, más certero será el análisis posterior de los datos disponibles, ya que permite una agrupación correcta por categorías.

$$Dato\ homogéneo = Dato\ específico \in Categorización\ homogénea$$

$$\%\ Homogeneidad\ Atributo = \frac{Datos\ homogéneos}{Datos\ totales}$$

$$n = total\ de\ atributos\ que\ califican\ para\ evaluar\ homogeneidad$$

$$\%\ Homogeneidad\ Total = \frac{Homog\ Atrib_1 + Homog\ Atrib_2 + Homog\ Atrib_3 + \ldots + Homog\ Atrib_n}{n}$$

Los errores que se obtengan como resultado de esta dimensión se deben subsanar mediante un **proceso de limpieza de datos.** Por tal motivo, conviene aplicar esta dimensión en dos oportunidades: una primera vez para conocer el estado de situación

de los datos de origen respecto a homogeneidad, y una segunda vez después del proceso de limpieza de datos (excepto en el caso de que la primera evaluación arroje una homogeneidad perfecta y no sea necesario implementar ningún proceso de limpieza de datos).

✅ **Ejemplos**

Imaginemos que se establece, en el modelo de negocio, que para el campo "día_semana" los registros posibles son **LUN, MAR, MIE, JUE, VIE, SAB, DOM.** Si algún registro utilizara el nombre completo (lunes, martes, miércoles, etc.), o cualquier otra alternativa que no sea una de las indicadas, ese registro no sería homogéneo.

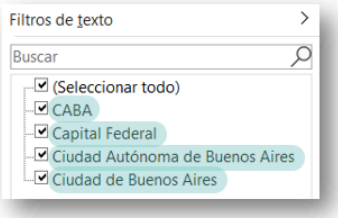
Para facilitar otro ejemplo, supongamos que se establece en la guía de datos que para el campo "tipo_documento" los registros posibles son **pasaporte, dni, libreta_cívica, libreta_enrolamiento**. Si algún registro fuera *libretacivica,* todo junto, o *libreta_de_enrolamiento* con el "de" adicional, o cualquier otra alternativa que no sea una de las indicadas, tampoco cumpliría con el requisito de homogeneidad.

En la siguiente tabla es posible ver que los datos del campo "*ciudad*" y "*país*" no son homogéneos en su totalidad.

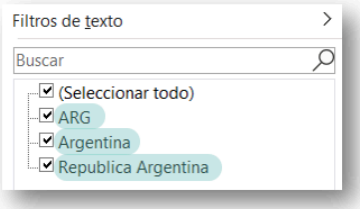
Homogeneidad

id	posta_digital	latitud	longitud	ciudad	país
134123	1	-346110	-583738	Capital Federal	Argentina
134124	2	-345886	-583974	Ciudad Autónoma de Buenos Aires	Argentina
134125	3	-346141	-584056	CABA	Argentina
134126	4	-346372	-583828	Ciudad de Buenos Aires	Argentina
134127	5	-346194	-584210	Ciudad Autónoma de Buenos Aires	Argentina
134128	6	-346192	-584333	Ciudad Autónoma de Buenos Aires	Argentina
134129	7	-346295	-584510	Ciudad Autónoma de Buenos Aires	ARG
134130	8	-346824	-584747	Ciudad Autónoma de Buenos Aires	Republica Argentina
134131	9	-346441	-584944	Ciudad Autónoma de Buenos Aires	Argentina
134132	10	-346321	-585006	Ciudad Autónoma de Buenos Aires	Argentina

4 formas diferentes de denominar a la **Ciudad de Buenos Aires**



3 formas diferentes de denominar **Argentina**



Estos errores tienen implicancias negativas en una etapa de análisis y visualización de datos, ya que pueden generar confusiones a la hora de seleccionar una categoría a filtrar y conocer el comportamiento de los datos respecto a ella.

Clasificación heterogénea

ciudad	país
Capital Federal	Argentina
Ciudad Autónoma de Buenos Aires	Argentina
CABA	Argentina
Ciudad de Buenos Aires	Argentina
Ciudad Autónoma de Buenos Aires	Argentina
Ciudad Autónoma de Buenos Aires	Argentina
Ciudad Autónoma de Buenos Aires	ARG
Ciudad Autónoma de Buenos Aires	Republica Argentina
Ciudad Autónoma de Buenos Aires	Argentina
Ciudad Autónoma de Buenos Aires	Argentina

4 formas diferentes de denominar a la **Ciudad de Buenos Aires**

3 formas diferentes de denominar **Argentina**

Clasificación homogénea

ciudad	país
CABA	Argentina
CABA	Argentina
CABA	Argentina
CABA	Argentina
CABA	Argentina
CABA	Argentina
CABA	Argentina
CABA	Argentina
CABA	Argentina
CABA	Argentina

Formas univoca de clasificar **ciudad**

Forma univoca de clasificar **país**

4.3.2.1.7. Correctitud

Para la dimensión de **correctitud** se analizan los atributos de tipo alfanumérico (string/varchar/texto) y numéricos a los cuales se les puedan exigir ciertas *condiciones tipográficas, de formato o lógicas* según las reglas del modelo de negocios y guía de datos.

🔍 **Fuentes de origen de datos**

La correctitud aplica a todas las fuentes de datos.

💡 **Resultados esperados**

Se debe obtener un porcentaje de los atributos analizables, que refleje en qué medida cumplen con esta condición. Cuanto mayor sea el porcentaje obtenido, podemos considerar que los datos se encuentran alineados con una mayor calidad.

$$\%\ Correctitud\ Atributo\ =\ \frac{Datos\ correctos}{Datos\ totales}$$

$$n\ =\ total\ de\ atributos\ que\ calif\,ican\ para\ evaluar\ correctitud$$

$$\%\ \ Correctitud\ Total\ =\ \frac{Correct\ Atrib_{1}+\ Correct\ Atrib_{2}+\ Correct\ Atrib_{3}\ +\ \ldots\ +\ Correct\ Atrib_{n}}{n}$$

Los errores que se obtengan como resultado de esta dimensión se deben subsanar mediante un proceso de limpieza de datos. Por tal motivo, y al igual que la homogeneidad, conviene aplicar esta dimensión en dos oportunidades: la primera vez para conocer el estado de situación de los datos de origen respecto a correctitud, y la segunda vez después del proceso de limpieza de datos (excepto en el caso de que la primera evaluación arroje una correctitud elevada y no sea necesario implementar ningún proceso de limpieza de datos).

✅ **Ejemplos**

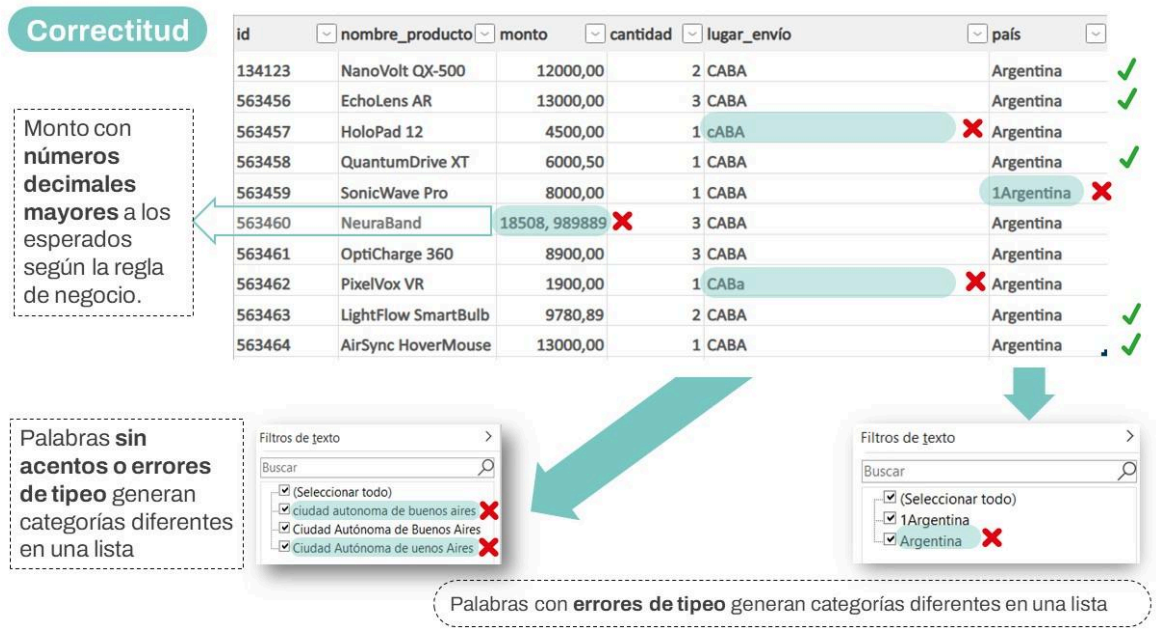
Supongamos que contamos con un atributo llamado “monto”, el cual es de tipo decimal (38,10). Si bien en su definición de formato este atributo puede aceptar registros de cifras con 10 decimales, no sería lógico que tuvieran más de 2 decimales de acuerdo con la denominación del peso argentino.

Otro ejemplo que no cumpliría con la correctitud sería tener un campo denominado “número_documento” que registre pasaportes con menos de 7 o más de 12 caracteres.

Tampoco sería lógico tener un atributo llamado “año” que, siendo de tipo decimal (10,0), acepte números menores o mayores a 4 cifras.

Para cerrar con un ejemplo de tipo string, atributos como “nombre” o “apellido” no deberían aceptar números o símbolos como **#$%&/()**, así como tampoco las palabras TEST, PRUEBA o ABC. En el caso de los apellidos sí debieran aceptar los registros equivalentes a “Testa” y “Malatesta”, dado que son apellidos reales.

En la tabla a continuación, vemos que la columna “*país*” contiene un registro con números, cuando debería contener solamente letras. También vemos que en la columna “lugar_envío” existen registros, señalados con color, que no cumplen con las condiciones de formato establecidas.



En ocasiones, cuando se analiza la correctitud de atributos categorizables, pueden existir similitudes con el análisis de homogeneidad. Ejecutar en primera instancia el análisis de homogeneidad, facilita que los atributos lleguen más limpios al análisis de correctitud. Recordemos que la correctitud es un análisis más amplio ya que evalúa todos los atributos (no sólo los categorizables) y lo hace en torno a varias cuestiones tipográficas, de formato e incluso lógicas.

4.3.2.1.8. Coherencia

El análisis de la **coherencia** implica conocer si los datos respetan una lógica interna. Es una dimensión cuyo análisis conlleva necesariamente un tiempo de procesamiento considerable, ya que en ocasiones requiere analizar distintos campos e incluso distintas tablas en simultáneo.

Por este motivo, si bien será mencionado como parte de la secuencia del análisis de calidad de datos, su aplicación puede ser opcional. Para decidir si conviene evaluar la coherencia de una fuente de origen, es necesario conocer en detalle las reglas de negocio y sopesar la relación costo-beneficio, indagando críticamente si no se han obtenido hasta el momento suficientes conclusiones sobre la calidad de los datos al ejecutar el análisis de las dimensiones previas.

🔍 **Fuentes de origen de datos**

La coherencia aplica a todas las fuentes de datos.

💡 **Resultados esperados**

Se debe obtener un porcentaje de los atributos analizables que refleje en qué medida cumplen con esta condición. Una vez más, cuanto mayor sea el porcentaje obtenido, podemos considerar que los datos se encuentran alineados con una mayor calidad.

$$\%\ Coherencia\ Atributo = \frac{Datos\ coherentes}{Datos\ totales}$$

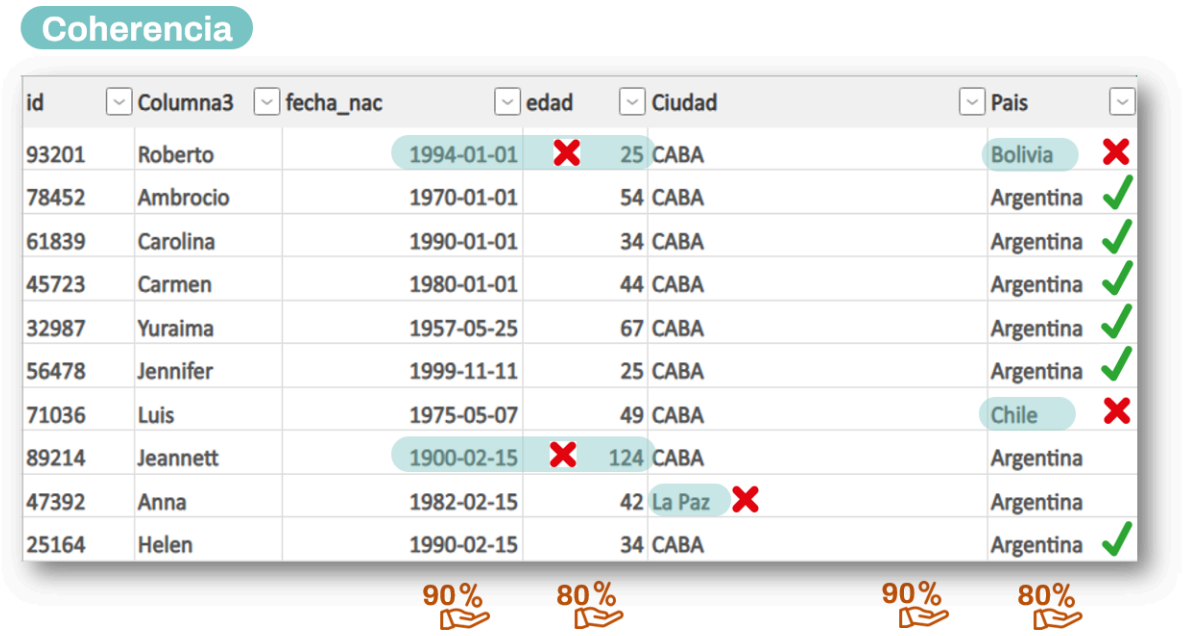
$$n = total\ de\ atributos\ que\ calif ican\ para\ evaluar\ coherencia$$

$$\%\ Coherencia\ Total = \frac{Coher\ Atrib_1 + Coher\ Atrib_2 + Coher\ Atrib_3 + \ldots + Coher\ Atrib_n}{n}$$

✅ **Ejemplos**

Teniendo en cuenta que la edad máxima alcanzada en Argentina es de una mujer de 115 años, no sería coherente que varias personas acusen dicha edad (o superior) en la base de datos. Por el mismo motivo, tampoco sería coherente que una persona haya nacido el 01/01/1900 o una fecha anterior, ni sería coherente que una persona tenga registrada una fecha de nacimiento igual o mayor a la fecha actual (fecha de consulta de la base).

Otro ejemplo podría estar dado por un registro en el que el valor monetario de un producto fuera negativo, o que la fecha de modificación de un suceso sea anterior a la fecha de alta o de registro de ese mismo suceso. Estas consideraciones van de la mano de las reglas del modelo de negocio sobre el que estemos trabajando: si tales reglas definen que la **fecha_modificación** de una contraseña solo puede ser posterior a la **fecha_registro** de esa misma contraseña, esa relación se tiene que mantener coherente en toda la base de datos analizada. Si en algún registro no se cumpliera, entonces estaríamos frente a un caso de incoherencia de los datos.



4.3.2.1.9. Conformidad

La **conformidad** evalúa si la totalidad de los datos y la metadata se encuadra dentro de los estándares del **Diccionario de Datos y Metadatos** del Gobierno de la Ciudad de Buenos Aires, detallado en el Anexo I de la presente guía.

🔍 **Fuentes de origen de datos**

La conformidad aplica a todas las fuentes de datos.

💡 **Resultados esperados**

Como resultado de este análisis se espera obtener no un porcentaje de conformidad sino una respuesta dicotómica: **SÍ o NO**.

4.3.2.1.10. Tabla de resumen con el ámbito de aplicación de cada dimensión

Dimensión	Fuente de origen a la que aplica
Relevancia	La metadata, el DER lógico y las definiciones ontológicas.
Completitud	Los atributos que el modelo de negocios define, los cuales deben estar completos.
Exactitud	Los atributos varchar, que, por lógica, deberían ser fechas, números o booleanos.
Unicidad	Los identificadores/PK y los atributos que, por lógica, deberían ser únicos.
Validez	Los atributos que sean factibles de validar a través de una fuente de validación externa.
Homogeneidad	Los atributos que sean categorizables.
Correctitud	Los atributos varchar o numéricos, a los que se les pueda exigir ciertas condiciones de formato, tipográficas o lógicas.
Coherencia	Los atributos de cualquier tipo, de los que se conozcan condiciones específicas que deban cumplir.
Conformidad	Todos los datos y la metadata.

4.3.2.2. Ponderación de calidad de datos

Como se mencionó a lo largo de esta guía, existen coincidencias en la evaluación de dimensiones de calidad de datos respecto a los resultados esperados, interpretados en un porcentaje de cumplimiento de la regla aplicada.

Estos resultados pueden resumirse de la siguiente manera:

$$Caso\ exitoso = Dato\ específico \in Regla\ de\ dimensión\ de\ calidad\ de\ datos$$

$$\%\ Atributo\ evaluable = \frac{Casos\ exitosos}{Casos\ totales}$$

$$n = cantidad\ de\ atributos\ evaluables$$

$$\%\ Dimensión\ de\ calidad = \frac{Atrib\ Evaluab_1 + Atrib\ Evaluab_2 + Atrib\ Evaluab_3 + \ldots + Atrib\ Evaluab_n}{n}$$

Asimismo, si se quisiera tomar una ponderación total de calidad de datos, se podría hacer de manera equitativa a las **N** dimensiones que se adopten en el proceso, es decir:

$$N = cantidad\ de\ dimensiones\ evaluadas$$

$$\%\ Calidad\ Total = \frac{\%\ Dim_1 + \%\ Dim_2 + \ldots + \%\ Dim_N}{N}$$

También podría adoptarse un criterio de ponderación especial, otorgándole un criterio de más importancia a ciertas dimensiones frente a otras, por ejemplo:

Dimensión	Ponderación
Relevancia	20 %
Completitud	20 %
Exactitud	20 %
Unicidad	10 %
Validez	10 %
Homogeneidad	10 %
Correctitud	5 %
Conformidad	5 %
8 dimensiones	**100 %**

Entonces:

$$\%\ Calidad\ Total = 0.2\ \%\ Relev + 0.2\ \%\ Comple + 0.2\ \%\ Exact + 0.1\ \%\ Unic + 0.1\ \%\ Valid + 0.1\ \%\ Homog + 0.05\ \%\ Correct + 0.05\ \%\ Conf\ or$$

La estrategia por adoptar será susceptible a las reglas de modelo de negocio que aplique a la fuente de origen, así como también el conocimiento que se tenga sobre el conjunto de datos.

4.3.2.3. Estrategias de estandarización y limpieza de datos erróneos.

La **limpieza y estandarización de dato**s son pasos fundamentales en el proceso de análisis de datos ya que garantizan la calidad, la precisión y la utilidad de la información. Hasta ahora, nos limitamos en esta guía a buscar herramientas de diagnóstico que nos podrían dar una noción general sobre la calidad de datos de diversas fuentes de origen. Adoptar estrategias de limpieza y estandarización de datos es el tratamiento necesario para mejorar las puntuaciones que afectan a las diversas dimensiones de calidad.

A continuación, se describen algunas estrategias clave en este proceso.

4.3.2.3.1. Identificación y eliminación de duplicados

Uno de los problemas más comunes en conjuntos de datos es la **presencia de registros duplicados.** Para limpiarlos, es necesario identificar las claves que pueden indicar duplicación (como un ID único) y luego aplicar técnicas para eliminar o consolidar estos registros.

4.3.2.3.2. Tratamiento de valores faltantes

En muchos casos, los conjuntos de datos contienen valores faltantes. Existen diversas estrategias para manejar estos casos, como:

- **Eliminación de registros**: si un registro tiene demasiados valores faltantes y no es representativo, puede eliminarse.
- **Imputación**: se pueden imputar valores a partir de la media, mediana o moda de la columna, o incluso utilizando modelos de machine learning.

4.3.2.3.3. Estandarización de formatos

Es crucial que todo el conjunto de datos siga un **formato uniforme.** Esto incluye la estandarización de fechas, direcciones, nombres y otros campos críticos. Por ejemplo, transformar todas las fechas en un formato específico (AAAA-MM-DD) facilitará el análisis y evitará confusiones.

4.3.2.3.4. Validación de datos

Implementar **reglas de validación** puede ayudar a detectar errores comunes de entrada de datos, como campos fuera de rango, caracteres inesperados o formatos incorrectos. Esto se puede lograr a través de scripts automatizados que revisen la integridad de los datos o incluyendo reglas de validación en el momento de la captura del dato.

4.3.2.3.5. Normalización de datos

Este proceso implica **ajustar los valores de diferentes escalas a una escala común**, lo cual es particularmente útil en análisis estadístico y aprendizaje automático. Por ejemplo, es posible escalar valores de ingresos para que todos estén en una misma proporción.

4.3.2.3.6. Documentación y gobernanza de datos

Documentar las decisiones tomadas durante los procesos de limpieza y estandarización es vital. Esto no solo ayuda en la replicabilidad del proceso, sino que también ofrece claridad sobre la calidad de los datos a otros usuarios o stakeholders.

5. ROLES Y RESPONSABILIDADES

5.1. Importancia de definir roles y responsabilidades en la gestión de los datos

En el marco de una gestión eficiente y eficaz de los datos, es esencial establecer roles definidos dentro de cada unidad organizativa del Gobierno de la Ciudad Autónoma de Buenos Aires para asegurar una **gobernanza adecuada** y un **manejo ordenado de la información.** La asignación de roles y responsabilidades no solo facilita la **transparencia en la gestión de datos**, sino que también optimiza la organización en la transferencia de datos y el manejo de solicitudes.

Es importante destacar que hablamos de **unidad organizativa** ya que la misma puede definirse a nivel de Ministerio, Secretaría, Subsecretaría o lo que corresponda en función del peso que un área tenga en la gestión de los datos.

Cuando una persona usuaria solicita acceso a un activo de datos, es necesario realizar una revisión del requerimiento de información dentro del marco de la clasificación y transferencia de datos para asegurar la protección debida. Las secciones anteriores sirven de guía para que los siguientes roles estén informados sobre cómo deben compartir los datos y con quiénes podemos hacerlo, y a mantener o incrementar la calidad de los mismos desde la fuente donde se generan.

Las áreas del Gobierno de la Ciudad deberán reconocer, dentro de sus equipos de trabajo, los roles definidos en esta guía para desarrollar un modelo óptimo de gobernanza de datos y ordenar la transferencia de manera interna y externa.

5.2. Modelo organizacional de gobernanza de datos

El modelo propuesto para la gobernanza de datos en el Gobierno de la Ciudad de Buenos Aires es híbrido. La **Dirección General de Gobernanza de Datos**, **dependiente de la Secretaría de Innovación y Transformación Digital,** es la repartición responsable de la definición de políticas y lineamientos de gestión de datos que todas las áreas deben cumplir. Esta Dirección General establece el marco normativo y las mejores prácticas para garantizar la interoperabilidad y la protección de la información en todo el Gobierno de la Ciudad.

A su vez, de manera descentralizada, **cada unidad organizativa,** según su nivel (Ministerio, Secretaría, Subsecretaría, Dirección), **debe designar a responsables específicos para la implementación y cumplimiento de estas políticas dentro de su propio ámbito.** Estos responsables deben asegurarse de que los datos generados, procesados y compartidos en su área cumplan con las normas establecidas y los estándares de calidad.

Este modelo permite que la Dirección General de Gobernanza de Datos defina el enfoque global de la gestión de datos, mientras que las unidades organizativas asumen la responsabilidad operativa sobre sus propios datos, facilitando la ejecución diaria de las políticas. Así, se garantiza una coherencia estratégica a nivel de todo el Gobierno, pero con flexibilidad en la implementación por unidad.

5.3. Roles necesarios dentro de la organización

5.3.1. Persona responsable de la gobernanza de datos

La **persona responsable de gobernanza de datos**, rol ejercido por el Director General de Gobernanza de Datos, o el área que en algún futuro la reemplace, lidera la estrategia de datos y asegura la correcta implementación de las políticas de gestión de datos a nivel de todo el Gobierno de la Ciudad. Este rol es el responsable de:

- **Definir la estrategia de datos dentro del Gobierno de la Ciudad de Buenos Aires** y brindar asistencia a las distintas unidades organizativas, para que la gestión de datos esté alineada con los objetivos institucionales y que se utilicen de manera efectiva para la toma de decisiones.

- **Definir políticas y lineamientos** que establezcan cómo los datos deben ser gestionados, almacenados, compartidos y protegidos, siempre en cumplimiento con las leyes y regulaciones vigentes.

- **Definir estándares** para asegurar la calidad de los datos incluyendo aspectos fundamentales como la exactitud, consistencia, integridad y actualización de la información.

- **Crear y mantener actualizado el Diccionario de Datos y Metadatos del Gobierno de la Ciudad**, un recurso centralizado para la definición y estandarización de los términos y estructuras de datos utilizados en toda la organización.

- **Realizar auditorías periódicas** para asegurar el cumplimiento de los lineamientos.

- **Promover una cultura basada en datos,** proporcionando programas de capacitación sobre la gestión y uso de los datos, y fomentando el uso de herramientas innovadoras para mejorar la gestión, almacenamiento y explotación segura de datos.

- **Facilitar la interoperabilidad de los datos** entre diferentes áreas de gobierno, entidades privadas y otras jurisdicciones.

- **Impulsar la publicación de datos relevantes** en el Portal de Datos Abiertos de la Ciudad, Buenos Aires Data, para garantizar la transparencia y el acceso a la información pública.

5.3.2. Persona responsable de datos en las unidades organizativas

Cada unidad organizativa debe designar una **persona responsable de datos**, quien actuará como el principal punto de contacto con la Dirección General de Gobernanza de Datos. Esta persona también puede concentrar las responsabilidades de los otros roles operativos si el área no cuenta con la estructura recomendada. Este rol es el responsable de:

- Enlazar y coordinar la implementación de las políticas y lineamientos de gestión de datos propuestas por el responsable de la Gobernanza de Datos en su unidad.

- Asegurar que las solicitudes de acceso y uso de datos se gestionen correctamente y mediar en situaciones donde pueda haber desacuerdo.

- Fomentar el conocimiento y la adopción de los lineamientos de gobernanza de datos dentro de la unidad organizativa.

- Promover la apertura y el uso compartido de datos dentro de su organización.

La **persona responsable de datos** es, por lo tanto, el **intermediario estratégico** entre la Dirección General de Gobernanza de Datos y los equipos operativos dentro de su área. Es importante que el responsable de datos pueda identificar a las personas dueñas y custodios de las distintas bases de datos de la organización para resolver cualquier consulta o necesidad de gestión.

5.3.3. Roles por bases de datos

Cada base de datos gestionada en una unidad organizativa debe contar con los siguientes roles:

5.3.3.1. Persona dueña de negocio

La **persona dueña de negocio** es responsable de que los datos sean funcionalmente útiles y cumplan con los requerimientos organizacionales. Este rol asegura que los datos se gestionen y utilicen para cumplir con los objetivos del negocio. Sus responsabilidades incluyen:

- Asegurar la calidad, accesibilidad, publicación y mantenimiento de los datos de los cuales tiene responsabilidad, garantizando que estos cumplan con los estándares del negocio.
- Entender la funcionalidad que genera los datos y cómo éstos son utilizados en los procesos operacionales.
- Definir y mantener la documentación funcional y el catálogo de datos de cada base de datos para cumplir con lo dispuesto en los lineamientos de calidad de datos.
- Conocer y aplicar los lineamientos de clasificación, calidad y transferencia de datos, tomando decisiones que cumplan con las normativas vigentes sobre protección de datos, principalmente al autorizar accesos y publicar datos.
- Supervisar la evolución de los sistemas que generan y utilizan los datos para adaptarse a las necesidades del negocio y soporten los procesos operacionales.

5.3.3.2. Persona dueña técnica

La **persona dueña técnica** es responsable de la gestión técnica de los sistemas y bases que generan y/o almacenan datos. Su rol está enfocado en la implementación, mantenimiento y seguridad técnica de estos sistemas, asegurando que los datos sean producidos, adaptados y disponibilizados en el momento adecuado. Conoce tanto la funcionalidad como los aspectos técnicos de la información. Sus principales funciones incluyen:

- Supervisar la infraestructura tecnológica y la seguridad de la base de datos asignada, garantizando su operatividad y protección.

- Mantener la operatividad y disponibilidad de los datos.
- Abordar y solucionar problemas técnicos relacionados con la base de datos, minimizando cualquier impacto y cumpliendo con los estándares establecidos.
- Otorgar acceso a los datos que han sido autorizados previamente por la persona dueña de negocio.
- Coordinar con la persona dueña de negocio y otros equipos técnicos para asegurar la correcta integración de los datos y minimizar la duplicación de entidades.
- Minimizar el impacto de los cambios en la estructura de datos, siguiendo los procesos de gestión del cambio, e informar y documentar cualquier modificación relevante.
- Desarrollar y actualizar catálogos y documentación siguiendo los estándares del Diccionario de Datos y Metadatos, y apoyar a la persona dueña de negocio en la definición de términos de negocio apropiados.
- Implementar y cumplir con los estándares de calidad y arquitectura de datos definidos por la persona responsable de la gobernanza de datos.
- Gestionar altas, bajas y modificaciones de accesos otorgados a los miembros de su área, en coordinación con la persona dueña de negocio.

5.3.3.3. Persona custodia

Los **custodios**, que pueden ser **custodios de negocio** o **custodios técnicos**, son las personas expertas en la materia, que realizan tareas operativas relacionadas con el manejo de los datos. Aunque no tienen responsabilidad directa en la toma de decisiones sobre los datos, sus funciones son clave para garantizar el buen funcionamiento del ciclo de vida de los datos. Sus tareas incluyen:

- Ejecutar los procesos definidos por las personas dueñas de negocio y técnicas, ya sea para una integración, transformación, explotación y consumo.

- Monitorear la calidad de los datos y reportar incidentes detectados.

- Asegurar que los datos se transformen y utilicen de manera correcta según las reglas establecidas.

- Documentar los procedimientos técnicos y funcionales para su correcta utilización, y mantener actualizado el catálogo de datos de cada base de datos.

- Actuar como experto en la materia en proporcionar consejos y soluciones en lo que respecta a los problemas de calidad de datos.

5.3.4. Persona usuaria de datos

Aunque no se considera un rol específico dentro de la gobernanza de datos, la **persona usuaria de datos** es fundamental en el ciclo de vida de los mismos. Es la principal destinataria, beneficiaria y consumidora de datos, y su grado de satisfacción es un indicador clave de la madurez del sistema de gobernanza de datos.

Además, una persona consumidora de datos puede ocupar cualquier otro perfil dentro de los roles de gobernanza, como dueño de negocio, dueño técnico o custodio e incluso responsable de datos en su unidad organizativa.

Los datos que recibe deben ser confiables, interoperables y accesibles para su uso y explotación autogestionada. Esto es importante para asegurar que el manejo de datos permita la toma de decisiones basadas en evidencia y optimice los procesos tanto en el ámbito público como privado.

Sus principales responsabilidades son:

- Ayudar a fomentar una cultura de datos basada en evidencia, promoviendo el uso eficaz y la toma de decisiones fundamentadas en la información disponible.

- Gestionar los datos de manera responsable y ética, asegurando su uso adecuado en concordancia con los principios de protección de datos.

- Fomentar internamente la autogestión del acceso, uso y explotación de los datos, promoviendo una mayor autonomía y eficiencia en su manejo.

- Colaborar con las personas dueñas técnicas y de negocio para evaluar y validar los requerimientos de datos, asegurando que los datos disponibles

sean compatibles con las necesidades del negocio o los objetivos establecidos.

- Informar sobre cualquier necesidad adicional de datos o requerimientos de reportes, contribuyendo a mejorar la calidad y utilidad de los datos gestionados.

5.4. Compatibilidad de roles

Es fundamental que cada unidad organizativa cuente con roles bien definidos para asegurar una gestión de datos eficiente y ordenada. No obstante, en ciertos casos, una misma persona puede asumir múltiples roles, siempre que exista compatibilidad entre ellos y que cuente con las competencias necesarias para cada función. Sin embargo, hay roles cuya incompatibilidad es esencial para preservar la independencia y efectividad en la gestión de los datos.

La **persona responsable de la gobernanza de datos**, que actúa a nivel organizacional y no dentro de una unidad específica, es única dentro del marco general del gobierno y **no puede asumir ningún otro rol.** Su imparcialidad y enfoque estratégico requieren que no se mezcle con funciones operativas dentro de las unidades.

La **persona responsable de datos** en cada unidad organizativa **puede asumir otros roles**, como el de persona dueña de negocio o técnica, especialmente en áreas con menos recursos o estructuras más pequeñas. Sin embargo, es vital garantizar que no haya conflictos de intereses y que todas las normativas de calidad y seguridad en la gestión de los datos se respeten y ejecuten de manera adecuada.

Cada base de datos debe contar con una **persona dueña de negocio y dueña técnica**. Aunque una misma persona puede ser responsable de varias bases de datos, es fundamental que mantenga el control y la supervisión necesarios para cumplir con sus responsabilidades.

5.5. Procedimiento para altas, bajas y modificaciones (ABM)

La **persona responsable de datos** de cada unidad organizativa debe ser designado mediante una Comunicación Oficial (CCOO) de la unidad organizativa, dirigida a la persona responsable de la gobernanza de datos, es decir, a la autoridad de la Dirección General de Gobernanza de Datos (DGGDA), o el área que en algún futuro la reemplace. Esta

designación formal asegura que la persona indicada asuma el rol con las responsabilidades y autoridad correspondientes para la gestión de datos dentro de su área.

Una vez designada, la persona responsable de datos es quien se encarga de mantener actualizada la información relacionada con el ABM (Altas, Bajas y Modificaciones) de las personas dueñas técnicas y dueñas de negocio. Para ello, se utiliza el **Registro de Roles de Gobierno de Datos,** almacenado en el Sistema de Administración Documental Electrónica (SADE), dentro del módulo Registro de Legajo Multipropósito (RLM), un sistema que permite el control centralizado de los responsables de cada base de datos y los roles asociados.

Es de gran importancia que el **Registro** se mantenga al día, ya que **proporciona la información necesaria para identificar a las personas correctas con quienes contactar** en caso de dudas, autorizaciones o cambios en la gestión de datos. La persona responsable de datos debe asegurarse de que cualquier modificación en los roles de dueño técnico o dueño de negocio se informe inmediatamente y se registre en el sistema.

Además, para reforzar el compromiso con la protección de los datos, es obligatorio que todas las personas que gestionan o manipulan datos firmen un acuerdo de confidencialidad, cuyo modelo se encuentra disponible en el Anexo II de este documento.

5.6. Ejemplos

✅ **Desarrollo de un tablero**

Una persona analista de negocio **(persona usuaria de datos)** necesita calcular ciertos indicadores de los cuales no posee los datos fuente. Un analista funcional **(persona custodia de negocio)** se encarga de definir y solicitar el tablero con los indicadores al área de BI. Un analista de BI **(persona custodia técnica**) realiza el análisis y, en conjunto con un desarrollador de ETL **(persona custodia técnica)**, realizan el desarrollo.

Para realizar esta tarea, las personas custodias técnicas obtienen el acceso a los datos por parte de la persona dueña del proceso que se va a medir **(persona dueña de negocio)**, y la información del modelo de datos de la persona analista técnica del área en cuestión **(persona dueña técnica)**.

Las personas responsables de datos de ambas áreas (consumidora y proveedora) son las personas facilitadoras del proceso, interactuando con los demás actores y generando consensos. Durante todo el desarrollo de este ciclo de vida del dato, los actores van recopilando y registrando metadatos en la plataforma de integración de datos que utilicen.

✅ **Publicación de un dataset geolocalizado en el portal Buenos Aires Data**

Una persona analista de negocio **(persona dueña de negocio)** de una determinada área de gobierno productora de datos solicita al equipo del portal Buenos Aires Data **(persona custodia de negocio)** la publicación de sus datasets geolocalizados.

El equipo de Buenos Aires Data pide al equipo de información geoespacial **(persona custodia técnica)** agregar la información solicitada a los datasets que brinda una persona analista técnica del área productora **(persona dueña técnica)**. Este equipo normaliza los datasets recibidos y agrega la información solicitada, para luego entregar el dataset resultante al equipo de Buenos Aires Data, quien se encarga de validar los datos con el área productora antes de disponibilizarlos para los vecinos y vecinas de la Ciudad de Buenos Aires **(personas usuarias de datos).**

Las personas responsables de datos de ambas áreas (consumidora y productora) son las facilitadoras del proceso, interactuando con los demás actores y generando consensos. Durante todo el desarrollo de este ciclo de vida del dato, los actores van recopilando y registrando metadatos en la plataforma de integración de datos.

✅ **Consumo de información cruda y procesada del sector privado**

Un analista de negocio **(persona usuaria de datos)** necesita realizar consultas, análisis y extracciones de características sobre datos fuente del sector privado (fuente externa al GCABA), a los cuales no tiene acceso. Una **persona custodia de negocio** de GCABA que recibe y analiza el alcance del requerimiento, en conjunto con la **persona custodia técnica** se ponen en contacto con la **persona dueña de negocio** y obtienen la autorización de acceso a los datos.

En esta gestión también se solicita la información del modelo de datos (si corresponde y hubiese) a la persona analista técnica del área privada en cuestión **(persona dueña técnica).** Un científico de datos **(persona custodia técnica)** realiza el análisis del requerimiento de la **persona usuaria de datos** y, a posteriori, aplica el desarrollo sobre la información a disponibilizar (cruda y procesada), de acuerdo a la especificación requerida, para que sea consumida.

La **persona responsable de datos del área** consumidora es la persona facilitadora con el sector privado, el cual es el productor del dato; y también es quien interactúa con los demás actores, generando consensos internos y externos a GCABA.

La **persona dueña técnica** brinda el acceso a la **persona usuaria de datos** para que consulte, analice y explote los datos. Durante todo el desarrollo de este ciclo de vida del dato, los actores van recopilando y registrando metadatos en la plataforma de integración de datos.

6. CONTACTO

Ante cualquier duda o comentario sobre este documento, podés escribirnos a **datosgcba@buenosaires.gob.ar**

ANEXO I

Diccionario de Datos y Metadatos

1. Introducción

El **Diccionario de Datos y Metadatos** es una herramienta esencial para la **gestión de información** en el **Gobierno de la Ciudad Autónoma de Buenos Aires (GCBA)**. Su **objetivo principal** es proporcionar una **definición estandarizada y detallada** de cada elemento de datos utilizado en las diversas áreas de gobierno, asegurando un modelo unificado para la denominación y estructuración de los datos. Esto garantiza una mayor **calidad de la información** y mejora la **interoperabilidad** entre sistemas y organizaciones.

Al establecer un **lenguaje común** para los datos, se facilita su **reutilización**, se **optimizan los procesos internos** y se mejora la **toma de decisiones** basadas en información precisa, contribuyendo así a una gestión pública más **eficiente y moderna**.

2. Beneficios del Diccionario de Datos y Metadatos

- **Interoperabilidad:** al estandarizar las definiciones de datos, se mejora la integración entre los sistemas y áreas del Gobierno de la Ciudad de Buenos Aires, facilitando la **comunicación interna** y **colaboración entre departamentos**. La coherencia en la terminología y manejo de la información evita malentendidos y errores.
- **Calidad de los datos:** el diccionario establece reglas y descripciones claras, ayudando a prevenir ambigüedades y errores en la manipulación de la información. Esto asegura que **todas las áreas** de gobierno trabajen con **definiciones consistentes**, garantizando que los datos sean precisos y confiables.
- **Soporte para la toma de decisiones:** al proporcionar acceso a definiciones precisas y bien estructuradas, el Diccionario de Datos y Metadatos mejora la capacidad del GCABA para tomar decisiones basadas en **información coherente** y **de alta calidad**. Esto impacta directamente en la capacidad del gobierno de diseñar **políticas públicas efectivas** y **responder a las necesidades ciudadanas con mayor eficiencia**.
- **Optimización del uso de datos:** la estandarización facilita la **reutilización de datos** a través de diversas aplicaciones y proyectos, lo que resulta en un mejor aprovechamiento de los recursos disponibles. Esto también contribuye a una mayor

eficiencia en la gestión de procesos internos y al **análisis de la información**, lo que refuerza la capacidad de respuesta del gobierno frente a desafíos administrativos y operativos.

3. Impacto en la transparencia y modernización

La implementación del Diccionario de Datos y Metadatos en el Gobierno de la Ciudad de Buenos Aires es un paso fundamental hacia la **transparencia y modernización de la gestión pública.** Esta herramienta no solo garantiza que el manejo de la información sea claro y coherente en todas las áreas de gobierno, sino que también fortalece la **confianza ciudadana** al asegurar que las decisiones gubernamentales se basen en **datos precisos** y **bien estructurados**.

Si bien los ciudadanos no interactúan directamente con este diccionario, su impacto es evidente en la **mejora de los servicios públicos**. Al tener un marco de referencia común para los datos, el Gobierno de la Ciudad puede **responder con mayor eficiencia** a las demandas de la ciudadanía, asegurando una **gestión más ágil y precisa**.

Además, se alienta a otras organizaciones, tanto públicas como privadas, a implementar sus propios diccionarios de datos para mejorar la **coherencia y precisión** en el manejo de la información, impactando positivamente en la **gestión interna** y en la **calidad de los servicios** ofrecidos.

4. Almacenamiento y actualización

El Diccionario de Datos y Metadatos se almacena en el Sistema de Administración Documental Electrónica (SADE), dentro del módulo Registro de Legajo Multipropósito (RLM).

Las definiciones de datos se actualizarán en tiempo real, manteniendo la información precisa y actualizada.

Ante la detección de una nueva necesidad de incorporación de datos, la Dirección General de Gobernanza de Datos, o el área que en un futuro la reemplace, analizará y realizará las adecuaciones sobre el Diccionario.

5. Atributos del Diccionario de Datos y Metadatos:

A continuación, se describen las propiedades que contempla cada uno de los registros del diccionario:

1. **ID del campo:** Identificador único del campo.
2. **Agrupación:** Clasificación general de los campos de datos (ej. geográfico, personal).
3. **Nombre descriptivo:** Nombre que describe el propósito y contenido del campo.
4. **Nombre del campo:** Identificador técnico utilizado en bases de datos.
5. **Descripción:** Explicación detallada del campo, indicando su función y tipo de datos.
6. **Fuente:** Origen del formato del dato, siguiendo estándares nacionales e internacionales.
7. **Clasificación de sensibilidad:** Nivel de sensibilidad que determina el grado de protección y acceso.
8. **Tipo de dato:** Tipo de información (texto, número, fecha).
9. **Formato:** Estructura específica del dato y su detalle.
10. **Observaciones:** Notas adicionales sobre restricciones o especificaciones.
11. **Ejemplos:** Ejemplos prácticos del dato que puede ingresar en el campo.
12. **Categoría:** Tipo de acceso al campo (libre o predefinido).
13. **Listas:** Opciones predefinidas disponibles, si aplica.

Para más información, consultar el **manual de uso del Diccionario de Datos y Metadatos del Gobierno de la Ciudad de Buenos Aires.**

ANEXO II

Modelo de Convenio de Confidencialidad

El modelo detallado a continuación obedece a criterios mínimos de cumplimiento, pudiendo la repartición actuante agregar contenido pertinente y adaptar a cada caso de uso particular el texto en cuestión:

ACUERDO DE CONFIDENCIALIDAD Y NO DIVULGACIÓN DE LA INFORMACIÓN

En virtud de los servicios prestados a la xxxxxxxx (en adelante, la Repartición), quien suscribe ..... (nombre y apellido), DNI Nº....., declaro conocer que los datos a los que tengo acceso en virtud de las tareas desempeñadas, se encuentran amparados bajo normas de confidencialidad indicadas a continuación.

El presente acuerdo se ajusta a lo mentado por la Ley de la Ciudad Autónoma de Buenos Aires N° 1.845 de Protección de Datos Personales, su Decreto Reglamentario Nº 725/2007, y toda aquella normativa concordante y relacionada, como también aquellas que en el futuro la sustituyan, amplíen, modifiquen y/o regulen. Al respecto, quien suscribe declara conocer el contenido de dichas normas.

Quien suscribe se obliga de forma irrevocable a no revelar, divulgar o transmitir, ceder, ni facilitar de cualquier forma o por cualquier medio, ya sea por acción u omisión, a toda persona humana o jurídica, y a no utilizar para su propio beneficio o de cualquier otra persona, humana o jurídica, cualquier información, dato, o documentación a la que acceda o sobre la que tenga conocimiento como consecuencia de la prestación del servicio referido en el párrafo anterior.

Dentro de esta obligación también se incluye cualquier información, dato o documentación de terceras partes o personas a la que hubiese accedido o tenido conocimiento en la prestación del referido servicio, y la administración de los datos almacenados o transferidos a la presente Repartición, o sobre los que la referida Repartición hubiese actuado de intermediaria. La enunciación precedente no tiene carácter taxativo. Ante la duda prevalecerá el carácter confidencial de la información, dato o documentación.

Quien suscribe declara conocer que la obligación de confidencialidad, reserva y no divulgación de la información se extiende a todas las personas integrantes de la Repartición, sin importar el modo de vinculación con ésta.

En consecuencia, quien suscribe, se compromete a realizar las acciones debajo previstas, sin perjuicio de aquellas otras obligaciones que le correspondan por ley. La siguiente enumeración no es taxativa:

1. Guardar la máxima reserva y secreto sobre la Información Confidencial en los términos del presente Acuerdo;
2. Utilizar la Información Confidencial únicamente para la prestación de los servicios prestados a la repartición. Todo otro uso distinto al aquí dispuesto deberá contar con el expreso consentimiento de autoridad competente.
3. No reproducir por ningún medio la Información Confidencial, excepto en la exacta medida en que ello resulte necesario para la prestación de los servicios a la repartición, y siempre que dicha reproducción no implique poner la Información Confidencial al alcance de terceros.
4. No divulgar, proporcionar, o revelar a terceros, en cualquier forma, la Información Confidencial.
5. Restituir toda la Información Confidencial al solo requerimiento de la Repartición.
6. Observar y adoptar cuantas medidas de seguridad sean necesarias para asegurar la confidencialidad, secreto e integridad de la Información Confidencial.
7. Observar y adoptar las medidas necesarias para garantizar la debida Protección de los Datos Personales.
8. Observar todas las políticas de seguridad implementadas por la Repartición con respecto a la Información Confidencial y la Agencia de Seguridad Informática (ASI).
9. A reconocer la propiedad intelectual y transferir, en favor de la Repartición, todo trabajo y/o proyecto; desarrollo; conocimiento; producido; desarrollado y/o proyectado, por quien suscribe en el marco y transcurso de la prestación de sus servicios.

Esta obligación de confidencialidad, reserva y no divulgación de la información subsistirá sin vencimiento de plazo, aún después de finalizada la prestación de los servicios a la Repartición, asumiendo la responsabilidad penal, administrativa y/o civil de los daños y perjuicios que por dolo y/o negligencia pudiera ocasionar la violación de los términos mencionados en el presente Acuerdo.

En....., a los.....días del mes de.....de 20.....-


Buenos
Aires
Ciudad
Vamos por más