



**Superintendencia
Financiera de Colombia**

Delegatura para Riesgo de
Lavado de Activos y
Financiación del Terrorismo

Guía para evaluar la calidad de datos de los clientes de las entidades vigiladas por la SFC

Marzo de 2025

CONTENIDO

1.	INTRODUCCIÓN	1
2.	DEFINICIONES.....	2
3.	MÉTRICAS	2
4.	GUÍA PARA EL SUPERVISOR	3
5.	INDICADOR DE CALIDAD DE DATOS	4
6.	COMANDOS DE REFERENCIA	6
7.	CONCLUSIONES Y RECOMENDACIONES	7

1. INTRODUCCIÓN

La calidad de los datos de los clientes es un aspecto importante para la adecuada gestión del riesgo de LAFT. Datos de calidad refleja la correcta implementación de las políticas de conocimiento de cliente, facilitan los procesos de debida diligencia, y ayudan a identificar mejor la exposición al riesgo.

De acuerdo con el marco conceptual para la evaluación del SARLAFT, para la Superintendencia Financiera de Colombia (SFC) es necesario evaluar y hacer seguimiento a la calidad de los datos de los clientes de las entidades vigiladas. De estas actividades, la SFC evaluará parte de la implementación de la gestión operativa que contribuye a los programas de prevención del riesgo de LAFT de las entidades vigiladas.

Para analizar la evolución y determinar la efectividad de las actividades que llevan a mejorar la calidad de los datos, la SFC debe realizar evaluaciones comparables en el tiempo y entre instituciones financieras. Dicho esto, el propósito de este documento es guiar al supervisor en la evaluación de la calidad de los datos de los clientes de las entidades vigiladas.

Esta guía toma elementos de la *Data Management Association (DAMA)*, organización que establece estándares para la gestión de datos, incluyendo criterios para evaluar su calidad, tales como: completitud, validez, unicidad, precisión, consistencia y oportunidad. Estas dimensiones son buenas prácticas que están en armonía con la Circular Básica Jurídica¹ e incorporan buenas prácticas que favorecen la gestión del riesgo de LAFT de las instituciones financieras.

Este documento está dividido en las siguientes secciones:

¹ Subnumeral 4.2.4.2.7, Capítulo IV, Título IV, Parte I, Circular Básica Jurídica 029 de 2014

1. Definiciones: describe cada una de las dimensiones de la calidad de los datos.
2. Métricas: presenta los indicadores cuantitativos utilizados por la Delegatura para riesgo de LAFT (DRLAFT) para evaluar la calidad de los datos.
3. Guía para el supervisor: proporciona instrucciones para calcular y verificar la calidad de los datos en las entidades vigiladas.
4. Indicador de calidad de datos: explica la metodología de agregación de las dimensiones y los criterios para evaluar la calidad de datos.
5. Comandos de referencia: presenta los comandos para construir los indicadores directamente desde las bases de datos.
6. Conclusiones y recomendaciones.

2. DEFINICIONES

Para concluir sobre la calidad de los datos de los clientes de las entidades vigiladas, el supervisor debe evaluar las siguientes dimensiones²:

Dimensión	Definición
Complejidad	Los datos se consideran completos cuando los campos requeridos para la prevención del riesgo de LAFT están presentes y disponibles para su uso.
Validez	Los datos se consideran válidos cuando los campos corresponden con el formato, tipo y rango esperado.
Unicidad	Mide el número registros duplicados. Los datos se consideran únicos si aparecen una sola vez en el conjunto de datos analizado.
Precisión	Datos precisos reflejarán la realidad. Por ejemplo, los ingresos declarados por el cliente cuando se vincula son ciertos.
Consistencia	Se refiere a la uniformidad de los valores de los datos en distintos conjuntos de datos. Por ejemplo, los ingresos registrados en la vinculación deben coincidir con los reportados en una solicitud de crédito.
Oportunidad	Datos oportunos son aquellos que están actualizados y disponibles en el momento en que se requieren.

3. MÉTRICAS

Para evaluar la calidad de los datos, el supervisor debe utilizar métricas específicas para cada dimensión. Éstas proporcionarán una base cuantitativa para identificar áreas de mejora, monitorear el progreso y asegurar que los datos cumplan con los estándares requeridos. A continuación se presentan las métricas de cada dimensión y variables de interés.

² las definiciones se basaron en las definiciones del Government Data Quality Hub del Reino Unido.
<https://www.gov.uk/government/news/meet-the-data-quality-dimensions>

Dimensión	Métricas
Compleitud	$x_i = 1 - \frac{N \text{ registros vacíos}_i}{\text{Total registros}_i}$, donde i es la variable de interés
Validez	$\text{Documento} = 1 - \frac{N \text{ registros con espacios}}{\text{Total registros}}$ $\text{Nombre} = 1 - \frac{N \text{ registros caracteres especiales}}{\text{Total registros}}$ $\text{Teléfono} = 1 - \frac{N \text{ registros dígitos} \neq 10}{\text{Total registros}}$ $\text{Fecha} = \frac{N \text{ registros formato fecha}}{\text{Total registros}}$ $V. \text{Financiera}_i = \frac{N \text{ registros numéricos}_i}{\text{Total registros}_i}$
Unicidad	$x_i = 1 - \frac{N \text{ registros duplicados}_i}{\text{Total registros}_i}$, donde i es la variable de interés
Precisión	$\text{Ingresos} = \left(\left(1 - \frac{N \text{ registros ingresos igual a cero}}{\text{Total registros}} \right) + \left(1 - \frac{N \text{ registros ingresos menores a egresos}}{\text{Total registros}} \right) \right) / 2$ $\text{Egresos} = \left(\left(1 - \frac{N \text{ registros egresos igual a cero}}{\text{Total registros}} \right) + \left(1 - \frac{N \text{ registros ingresos menores a egresos}}{\text{Total registros}} \right) \right) / 2$
Consistencia	$V. \text{Financiera}_i = 1 - \frac{\sum \text{Valores conjunto } A_i - \sum \text{Valores conjunto } B_i}{\sum \text{Valores conjunto } A_i}$
Oportunidad	$x_i = 1 - \sum_j \frac{N \text{ registros } [d_j + 30 > D_j]}{N_j}$ <p>donde i es la variable de interés d_j = fecha hoy – fecha actualización del campo* D_j es el número de días máximos por perfil de riesgo j es el perfil de riesgo del cliente</p>
* algunas entidades no registran la fecha de actualización de los campos y solo registran la última fecha de actualización de cualquier campo. En este caso, se deberá calcular indicador con esta única fecha y asignar a todos los campos medidos.	

4. GUÍA PARA EL SUPERVISOR

Para evaluar la calidad de los datos de los clientes de las entidades vigiladas, el supervisor debe:

- Compartir la guía con la entidad vigilada al momento de enviar la carta de apertura del proceso de supervisión.
- Calcular los indicadores únicamente con los clientes activos a la fecha de la evaluación.
- Medir, como mínimo, las variables que se presentan en la Tabla 1. El supervisor podrá evaluar más variables, siempre y cuando éstas estén definidas en las políticas de datos de cada entidad.

Tabla 1. Campos mínimos para la medición del indicador de calidad de datos

Campos	Dimensiones de la calidad					
	Compleitud	Validez	Unicidad	Precisión	Consistencia	Oportunidad
Número documento						
Nombre						
Fecha de nacimiento						
Fecha de vinculación						
Celular						
Correo electrónico						
Activos						
Pasivos						
Ingresos						
Egresos						

- Solicitar a las entidades vigiladas enviar los resultados de la medición de cada variable y dimensión en un archivo Excel.
- Solicitar a las entidades vigiladas registrar dos matrices separadas: una para clientes persona natural y otra para clientes persona jurídica.
- Realizar una prueba de recorrido para corroborar los resultados reportados en el archivo Excel. Dicha prueba consistirá en replicar la consulta sobre las tablas utilizadas por las entidades vigiladas y verificar los numeradores obtenidos por la entidad (ver Capítulo 6. Comandos de referencia).
- Calcular el indicador de calidad de datos.

5. INDICADOR DE CALIDAD DE DATOS

Para estimar el indicador de calidad de datos, el supervisor debe agregar las dimensiones medidas en cada una de las variables analizadas, teniendo en cuenta los siguientes lineamientos:

1. Reglas de aplicación:

- El cálculo del indicador debe realizarse exclusivamente sobre la base de datos de clientes activos.
- Para la dimensión de oportunidad, la metodología de cálculo dependerá de cómo la entidad registre la actualización de los datos en su sistema:

- Si la entidad registra actualizaciones por variable, el cálculo deberá realizarse por campo.
 - Si la entidad registra una única fecha de actualización independiente de la variable actualizada, el cálculo para cada campo utilizará la misma fecha de actualización. El supervisor deberá recomendar a la entidad vigilada registrar la fecha de actualización por campo.
- Para los campos resaltados en gris, no se medirá la dimensión correspondiente. Por ejemplo, la dimensión de precisión no aplica para el campo *nombre*.
 - Para los campos donde sí aplica la medición, la calidad de la dimensión se considerará adecuada cuando el resultado sea igual o superior al 90%.
 - Los campos que cumplan con esta condición tomarán el valor de 1. Los que no cumplan, tomarán el valor de 0.
 - Se debe realizar un cálculo independiente para personas naturales y personas jurídicas, asegurando que la evaluación refleje las características específicas de cada grupo.
 - Se recomienda la utilización de Microsoft Excel u otros sistemas automatizados para facilitar el cálculo y evitar errores manuales.

2. Cálculo del indicador:

El indicador se calculará así:

$$Calidad\ de\ datos = \frac{\sum_{ij} C_{ij}}{N}$$

Donde:

- $C_{ij} = 1$ si el campo i y la dimensión j es $\geq 90\%$
- $C_{ij} = 0$ si el campo i y la dimensión j es $< 90\%$
- N es el número total de campos evaluados en la matriz (los campos grises no deben ser incluidos).

3. Ejemplo práctico:

Supongamos que una entidad vigilada ha evaluado 10 campos, considerando cinco dimensiones de calidad para cada una. Los resultados obtenidos se documentan en una matriz donde cada celda representa el porcentaje de cumplimiento de la dimensión evaluada. Si el cumplimiento de una dimensión es igual o superior al 90%, se asigna un valor de 1; de lo contrario, se asigna un valor de 0.

Tabla 2. Ejemplo matriz para cálculo del indicador de calidad de datos

Campos	Completitud	Validez	Unicidad	Precisión	Consistencia	Oportunidad
Número documento	100,0%	100,0%	100,0%			37,2%
Nombre	100,0%	100,0%	100,0%			
Fecha de nacimiento	44,5%	100,0%				
Fecha de vinculación	100,0%	100,0%				

Campos	Completitud	Validez	Unicidad	Precisión	Consistencia	Oportunidad
Celular	91,8%	85,0%	99,4%			
Correo electrónico	99,2%	99,2%	98,9%			
Activos	44,5%	44,5%				
Pasivos	44,5%	44,5%				
Ingresos	44,5%	44,5%		42,2%		
Egresos	44,5%	44,5%		42,8%		

Procedimiento para calcular el indicador:

1. Valide el porcentaje de cumplimiento de cada dimensión evaluada y proceda a convertir los valores en porcentaje a una escala dicotómica de 1 a 0. Por ejemplo, la completitud del campo número de documento es del 100% por lo tanto tomará el valor de 1, mientras que la del campo fecha de nacimiento es del 44,5% por lo tanto tomará el valor de 0.
2. Sume todos los campos que tienen en el valor "1", en este caso da un valor de 14.
3. Cuente cuántos campos fueron evaluados (sin incluir las vacías). Para el ejemplo fueron 27 celdas evaluadas.
4. Divida la cantidad de valores "1" entre la cantidad total de celdas evaluadas y multiplica por 100. Calidad de datos= $(14/27) \times 100 = 52\%$.
5. Para este ejemplo, el resultado del indicador de calidad de datos para persona natural es de 52%.
6. Si el indicador es superior al 90% se considerará que la calidad de datos de los clientes de la entidad vigilada es adecuada. Si es inferior, se considerará que necesita mejorar o es débil.

6. COMANDOS DE REFERENCIA

Para corroborar el cálculo de los indicadores, el supervisor debe revisar que los comandos utilizados por las entidades vigiladas sean parecidos a los que se presentan a continuación. Es importante resaltar que los comandos de referencia están basados en MySQL, los cuales pueden ser adaptados según las características de los conjuntos de datos y los sistemas de información de las entidades vigiladas.

Dimensión	Comando
Completitud	<code>SELECT COUNT(*) - COUNT(column) FROM table</code>
Validez	<code>SELECT COUNT(*) FROM table WHERE LEN(documento) - LEN(REPLACE(documento, ' ', '')) > 0</code>
	<code>SELECT COUNT(*) FROM table</code>

Dimensión	Comando
	<pre>WHERE nombre_cliente_pn LIKE '%[^A-Za-z]%' SELECT COUNT(*) FROM table WHERE nombre_cliente_pj REGEXP '.*[^A-Za-z0-9].*' SELECT COUNT(teléfono) FROM table WHERE LEN(teléfono) <> 10 SELECT COUNT(*) FROM table WHERE email LIKE '%@%.%' SELECT COUNT(*) FROM table WHERE fecha > GETDATE OR fecha < '1940-01-01'</pre>
Unicidad	<pre>SELECT COUNT(*) FROM (SELECT column, COUNT(*) FROM table GROUP BY column HAVING COUNT(*) > 1)</pre>
Precisión	<pre>SELECT COUNT(*) FROM table WHERE column = 0</pre>
	<pre>SELECT COUNT(*) FROM table WHERE ingresos – egresos < 0</pre>
Consistencia	<pre>SELECT SUM(column) FROM tableA - SELECT SUM(column) FROM tableB</pre>
Oportunidad	<pre>SELECT COUNT(*) FROM table WHERE (GETDATE - column) + 30 > perfil de riesgo</pre>

7. CONCLUSIONES Y RECOMENDACIONES

Garantizar la calidad de los datos es fundamental para fortalecer la gestión del riesgo de LAFT. La aplicación de esta guía contribuirá a mejorar la evaluación y seguimiento de la calidad de los datos de los clientes. El supervisor debe tener en cuenta las siguientes recomendaciones al momento de interactuar con las entidades vigiladas:

1. Además del monitoreo periódico de la calidad de los datos de los clientes activos, implementar un sistema de alertas tempranas para identificar anomalías y desviaciones en tiempo real. Esto permitirá una acción correctiva inmediata, minimizando el riesgo de LAFT.
2. No solo sensibilizar a los equipos responsables, sino también proporcionar capacitación continua y especializada sobre las últimas tendencias y mejores

prácticas en calidad de datos y prevención de LAFT. Fomentar una cultura organizacional donde se promueva la colaboración entre áreas y se asignen claramente las responsabilidades en la gestión de datos.

3. Establecer estándares claros y consistentes para la captura, almacenamiento y procesamiento de datos. Definir roles y responsabilidades claras en la gobernanza de datos, asegurando la rendición de cuentas y la transparencia.
4. Implementar un proceso de evaluación continua para medir la efectividad de las actividades que buscan mejorar la calidad de datos y realizar ajustes según sea necesario.