

## Nuevo marco de calidad de datos de la EMA / HMA para la regulación de medicamentos

En diciembre de 2023, la EMA y la HMA publicaron conjuntamente la primera versión completa del documento sobre Marco de Calidad de Datos para la regulación de medicamentos en la EU. Este documento tiene como objetivo caracterizar, evaluar y garantizar la calidad de datos para el respaldo de la toma de decisiones regulatorias.



**MAR DIAZ**  
RESPONSABLE TÉCNICA  
VALIDACIONES TRESICAL  
LIFE SCIENCES

### Introducción y foco

El progreso en la digitalización y la tecnología de la información ha creado nuevas oportunidades que han contribuido a un panorama cada vez más complejo para la toma de decisiones regulatorias. Por ello, se está reconociendo una creciente necesidad de estandarización y la implementación de un marco que describa la calidad de datos. Este marco permitiría al regulador realizar evaluaciones regulatorias confiables sobre si los datos son apropiados para respaldar la toma de decisiones. Pero, aunque se dispone de nuevos tipos o estándares de datos, todavía son escasas las directrices o métodos para demostrar si dichos datos son adecuados para la toma de decisiones.

Por todo ello, se ha visto necesario por parte de la European Medicines Agency (EMA) y la Heads of Medicines Agencies (HMA) la elaboración de un documento Marco de Calidad de Datos (DQF, por sus siglas en inglés) que brinde orientación y guía para el logro de procedimientos de evaluación de calidad coherentes y consistentes. Este documento de 42 páginas, ha sido publicado en el mes de diciembre de 2023, tras ser actualizado tras los comentarios recibidos durante su versión preliminar de octubre de 2022.

Tal y como se reconoce en las recomendaciones del Grupo de Trabajo Conjunto de la HMA-EMA y el plan de trabajo del Grupo Directivo Conjunto de la HMA-EMA sobre Big Data, el establecimiento de un marco de la EU para la calidad y representatividad de los datos es un elemento crítico para aprovechar todo el potencial de los datos (big data) e impulsar las decisiones regulatorias.

Con el establecimiento de un marco de calidad de datos la EMA y la HMA pretenden:

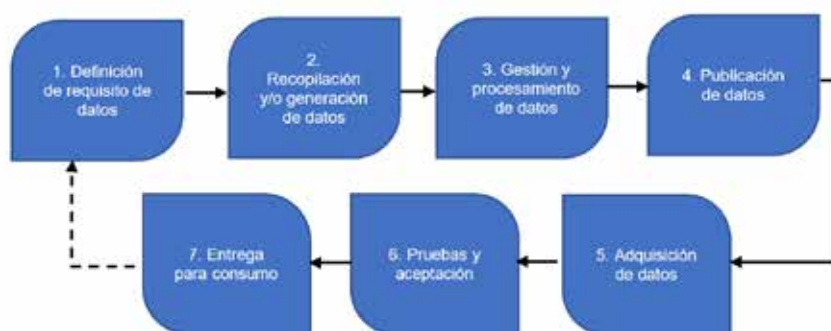


Figura 1: Flujo de trabajo típico de procesamiento de datos en el proceso de generación de evidencia.

- Mejorar la coherencia en la evaluación de la calidad de datos utilizados por los reguladores.
- Facilitar el desarrollo de un enfoque estandarizado para la calidad de datos en todas las fuentes de datos.
- Favorecer un uso más sistemático de los datos en la toma de decisiones regulatorias.
- Apoyar la confianza de los interesados en los datos que respaldan las decisiones regulatorias.

Este marco va alineado además con el objetivo principal del plan de trabajo del Grupo Directivo Conjunto de la HMA-EMA sobre Big Data que es el de desarrollar pautas para el logro de un proceso fortalecido para la cualificación de datos a través de asesoramiento científico, y promover en los Estados miembros la adopción de registros electrónicos de salud, registros, datos genómicos y la disponibilidad segura de los datos.

### Alcance y contenido

El documento es la primera publicación del Marco de Calidad de Datos de la EU (DQF) para la regulación de medicamentos y define principios y procedimientos de alto nivel que se aplican en todo el mandato regulatorio de la EMA. Por ello, brinda un análisis sobre las acciones y métricas de calidad de datos que

deben considerarse en diferentes casos de uso e introduce un modelo de madurez para guiar la evolución de la automatización para respaldar la toma de decisiones regulatorias basada en datos.

El objetivo es proporcionar un conjunto de definiciones, principios y directrices para su aplicación de forma coherente en cualquier fuente de datos con el fin de caracterizar, evaluar y asegurar la calidad del conocimiento para la toma de decisiones regulatorias. Este marco está destinado además a abarcar el uso primario y secundario de los datos, así como metadatos e información de respaldo.

Sobre las verificaciones de calidad de datos, estas ocurren en varios pasos a lo largo del proceso de generación de evidencia y pueden incluir ciclos de retroalimentación iterativos tal y como se indica la línea discontinua de la figura 1.

Por otra parte, dentro del alcance, el documento quiere hacer reseña del concepto de *dato*, que define como “cualquier activo de información que representa mediciones u observaciones y que pueden usarse para apoyar la toma de decisiones, directa o indirectamente a través del análisis”, así como del de *Calidad de datos* que define como “la idoneidad para el propósito de las necesidades de los usuarios en relación con la investigación en salud, la formulación de políticas y

la regulación, y que los datos reflejan la realidad que intentan representar.”

Se remarca por lo tanto que la calidad de datos debe ser evaluada desde el punto de vista de su idoneidad para satisfacer las necesidades de los usuarios.

Por otra parte, el documento está dirigido principalmente a la red reguladora de medicamentos de la EU, pero la relevancia del contenido puede ser de interés para una gama más amplia de partes interesadas, como titulares de autorizaciones de comercialización, titulares de fuentes de datos, investigadores y asociaciones de pacientes.

## Estructura

El documento DQF se compone de dos partes, que reflejan diferentes etapas en el proceso de especificación.

- La primera parte (marco general) está diseñada para proporcionar un enfoque coherente para la calidad de datos, que abarque una amplia gama de tipos de datos y sea extensible a casos de uso novedosos. Por ello, proporciona un terreno común sobre diferentes aspectos de calidad de datos que se aplican a diferentes tipos de datos y escenarios: definiciones, dimensiones de calidad de datos y ejemplos de métricas que cubren dichas dimensiones. Además, identifica patrones generales para la aplicabilidad de los procesos de calidad de datos y articula un conjunto de modelos de madurez diseñados para impulsar una mayor automatización de la toma de decisiones regulatorias sobre medicamentos basada en datos.
- La segunda parte (especialización del marco) especializa y eventualmente amplía recomendaciones genéricas para atender tipos de datos específicos o cuestiones regulatorias. Esta parte plantea la base para la derivación de directrices implementables reales, que deberán evolucionar a medida que los datos y las tecnologías cambien con el tiempo.

## Definiciones de interés

En el nuevo documento se establecen diferentes definiciones relevantes en lo concerniente a la Calidad de datos, que demuestran el detalle de conceptos y aclaraciones que quiere estar dotado el nuevo marco. A continuación, se recopilan a modo resumen las siguientes:

- *Apto para el propósito*: Poseer todas las características de calidad de datos necesarias para abordar un objetivo específico. El énfasis en la calidad de datos es garantizar que los datos sean adecuados para el propósito de realizar evaluaciones confiables sobre si los datos son aptos para la toma de decisiones y para respaldar la investigación en salud y la salud de la población.
- *Calidad de datos*: Antes ya se ha definido este concepto, que se ha descrito como la idoneidad para el propósito de las necesidades de los usuarios en relación con la investigación en salud, la formulación de políticas y la regulación, y que los datos reflejan la realidad que intentan representar. Pero aclarar que la calidad de datos es relativa a la pregunta de investigación y no aborda la cuestión de en qué nivel se mide la calidad, por ejemplo, el nivel de variable, o fuente de datos. Estos aspectos se abordan en los determinantes de calidad de datos y en las dimensiones de calidad de datos.
- *Determinantes de Calidad de datos*: Lo que contribuye a la calidad de datos o su caracterización se clasifica en este marco en tres categorías:
  - Determinantes Fundamentales: Aquello que afecta la calidad de un conjunto de datos y es externo al propio conjunto de datos (por ejemplo, sistemas y procesos que generan datos).
  - Determinantes Intrínsecos: Lo que se puede derivar en términos de calidad para un conjunto de datos en sí mismo, sin información sobre cómo llegaron los datos o su uso previsto.
- Determinantes específicos de cuestión: Lo que afecta la evaluación de la calidad de un conjunto de datos y depende estrictamente del uso previsto o real del conjunto de datos. Métricas de calidad de datos: Las métricas de calidad de datos pueden definirse como indicadores que se pueden aplicar a una fuente de datos para obtener evaluaciones de una o más dimensiones de calidad.
- *Dimensiones de Calidad de datos*: Los aspectos de la calidad de datos se dividen en grupos diferentes que responden a diferentes preguntas sobre los datos. Estas divisiones se llaman “dimensiones”. Este marco distingue cinco dimensiones que se pueden dividir aún más en subdimensiones:
  - Extensibilidad: La cantidad de datos en términos de cobertura y exhaustividad.
  - Coherencia: La consistencia y uniformidad de los datos a lo largo del tiempo y en diferentes partes del conjunto de datos.
  - Oportunidad: La actualidad y relevancia temporal de los datos.
  - Relevancia: La adecuación y pertinencia de los datos para el propósito previsto.
  - Confiabilidad: La precisión y la confianza en la fuente y el proceso de generación de datos.
- *Integridad de los datos*: La integridad de los datos se refiere al mantenimiento y garantía de la confiabilidad y consistencia de los datos a lo largo del tiempo, abarcando todo el ciclo de vida de los



Figura 2. Dimensiones de la calidad de datos

datos. Es un concepto más amplio que la calidad de datos, cuya definición precisa varía según el contexto, extendiéndose desde aspectos físicos hasta lógicos del procesamiento y almacenamiento de datos.

- **Marco de Calidad de datos:** Un Marco de Calidad de Datos proporciona un conjunto de definiciones, pautas y recomendaciones para evaluar y gobernar la calidad de datos. El marco presentado aquí aborda una amplia gama de fuentes de datos con el propósito de caracterizar, evaluar y asegurar la calidad de datos para la toma de decisiones regulatorias.
- **Metadatos:** Los metadatos se definen como “datos sobre datos”, proporcionando contexto sobre su propósito y generación. Es un conjunto de datos que describe e informa sobre otros datos, proporcionando contexto sobre su propósito, ubicación, variables clave, generación, formato y propiedad de un conjunto de datos. Los metadatos suelen ser publicados en catálogos de datos, que tienen el propósito de permitir que los datos sean descubiertos y verificados en cuanto a su idoneidad, sin revelar los propios datos.
  - Para la toma de decisiones regulatorias, los metadatos en general deben considerarse de manera similar a los datos. Es decir, si algún cambio en los metadatos requiriese una revisión de las conclusiones derivadas, entonces debería tratarse como datos desde la perspectiva de la calidad de datos. Por ello, en un contexto de calidad de datos, los metadatos no deben verse como limitados a métricas y descripciones resumidas de conjuntos de datos, sino que deben extenderse a la caracterización de fuentes, procesos y definiciones de elementos de datos.
- **Modelo de madurez:** Un modelo de madurez es un marco para evaluar los procesos, la tecnología y la estructura de una organización o función. Proporciona un enfoque estructurado para evaluar qué tan bien una organización o función gestiona sus procesos, políticas y prácticas de calidad de datos. El modelo define características clave en cada nivel para guiar la medición de la mejora continua en la calidad de datos a lo largo del tiempo.

- **Uso primario de datos:** El uso principal de datos de salud (electrónicos) es el procesamiento de datos de salud personales para la prestación de servicios de salud para evaluar, mantener o restaurar el estado de salud de la persona a la que pertenece, incluida la prescripción, dispensación y suministro de medicamentos y productos médicos, así como para los correspondientes servicios de seguridad social, administrativos o de reembolso.
- **Uso secundario de datos:** El uso secundario de datos de salud (electrónicos) se refiere al procesamiento de datos de salud con fines distintos al uso primario, como estadísticas nacionales, educación/docencia, investigación científica, etc. Los datos utilizados pueden incluir datos de salud personales recopilados inicialmente en el contexto del uso primario, pero también datos de salud electrónicos recopilados con el propósito de uso secundario.

Es importante remarcar que el DQF limita su alcance a los determinantes de la calidad de datos que son relevantes para la toma de decisiones regulatorias.

## Consideraciones en la implementación del DQF

Por otra parte, el nuevo documento marco proporciona un conjunto de observaciones y recomendaciones guía en la implementación del DQF para el logro de mayores niveles de madurez. El documento destaca las siguientes:

- **Asegurar la calidad en origen:** Al diseñar procesos de recopilación y generación de datos, los aspectos que afectan a la calidad de datos deben abordarse de forma inicial ya que cuanto más se alejan los datos del contexto original, más difícil resulta corregir los problemas. Este asunto es particularmente relevante en el caso de los metadatos, ya que el conocimiento del contexto de generación de datos está presente solo en el momento de la generación.
- **Disponibilidad de datos maestros (MDM) y datos de referencia:** La disponibilidad de MDM y datos de referencia tiene un impacto directo en la calidad de datos. A menudo es un requisito previo para la coherencia de los datos e incluso puede afectar la confiabilidad en algunos escenarios de producción de datos,

ya que la información desconectada puede generar información errónea.

- **Alineación del SGC:** Un Sistema de Gestión de Calidad es un enfoque formalizado adoptado por una organización que documenta procesos, procedimientos y responsabilidades para lograr políticas y objetivos de calidad (por ejemplo, Good Clinical Practices [GCP], Good Laboratory Practices [GLP] or Good Manufacturing Practice [GMP]). A su vez, estándares como la familia ISO 9000 definen el SGC en todas las industrias, pero se han desarrollado SGC más específicos para industrias o productos específicos. Dentro del sector Life Sciences podemos resaltar los siguientes estándares:
  - ISO 14155 y Directiva de la UE 2001/20/EC para GCP (datos de ensayos clínicos)
  - Quality System Regulation (QSR) e ISO 13485 para dispositivos médicos.
  - Directiva de la UE 2004/9/EC y 2004/10/EC para GLP (datos de investigación de laboratorio).
  - ISO 15189 e ISO 17025 (datos de laboratorios clínicos).
- **Controles sobre los sistemas informáticos:** Ante el actual marco de digitalización, los sistemas informáticos utilizados para crear, modificar, mantener, archivar, recuperar o transmitir datos son un elemento crucial a tener en cuenta cuando hablamos de calidad de datos. Debemos por ello, asegurar que disponemos de sistemas informáticos efectivos, confiables y de alta calidad. La validación del sistema informático garantiza que el software se implementa adecuadamente y que existan los controles de proceso necesarios para usarlo de acuerdo a especificaciones, incluyendo la documentación asociada, el control de acceso, la gestión de proveedores y las auditorías. Por otro lado, debemos contar con un ciclo de vida de desarrollo de software que incluya un sistema de garantía de calidad del software para garantizar su diseño, desarrollo y verificación cuando el dato del SW es el propio producto (ej.: software de dispositivo médico).
- **El papel de los estándares ISO de la industria:** La Organización Internacional de Normalización (ISO) ha elaborado normas que proporcionan marcos para



Figura 3: Atributos ALCOA ++

la implementación de diversos aspectos de la gestión de datos, que se prueban en el campo y para las cuales se establecen plataformas, servicios de apoyo y organismos de certificación. Estos estándares a menudo se desarrollan para la implementación de industrias donde no se aplica la toma de decisiones regulatorias de la EMA.

- ISO 9000: Describe los estándares para los sistemas de gestión de calidad en todos los niveles de una organización. Se podría considerar la adopción de esta norma si no se aplica ningún SGC específico de la industria.
- ISO 8000: Describe los estándares de gestión de calidad de datos de una organización sometidos a un ciclo de mejora continua. La ISO 8000-150 especifica los principios fundamentales de la gestión de la calidad de datos maestros, centrada en los procesos, y los requisitos para la implementación, el intercambio de datos y la procedencia.
- ISO 25012: Define un modelo general de calidad de datos para datos retenidos en un formato estructurado dentro de un sistema informático, típico de los datos considerados para la toma de decisiones regulatorias. Proporciona un marco para establecer requisitos de calidad de datos, medidas de calidad de datos y un plan para realizar evaluaciones de calidad de datos. En un modelo de Calidad de Datos se establecen las 15 características de Calidad de Datos que se deben tener en cuenta a la hora de evaluar las propiedades de un producto de datos determinado (exactitud, completitud, con-

sistencia, credibilidad, actualidad, accesibilidad, conformidad, confidencialidad, eficiencia, precisión, trazabilidad, comprensibilidad, disponibilidad, portabilidad y recuperabilidad).

- ISO 13485: Es la norma que especifica los requisitos de un sistema de gestión de la calidad para que una organización diseñe, construya y obtenga autorización para dispositivos médicos que cumplan sistemáticamente con los requisitos regulatorios y del cliente.
- **Consideraciones sobre integridad de datos:** ALCOA ++ es un marco para la integridad de los datos utilizado en toda la industria farmacéutica. Postula un conjunto de atributos que deben cumplir los datos y su documentación para que éstos sean Atribuibles, Legibles, Contemporáneos, Originales, y Acertados (ALCOA). El ++ se refiere a los siguientes atributos: Completo, Consistente, Duradero, Disponible, Trazable. Siguiendo esta premisa, debemos centrarnos en los determinantes fundamentales que afectan principalmente las dimensiones de Confiabilidad de los datos, pero también en las de Extensión, Coherencia y Oportunidad.
- **Controles sobre la Calidad de datos:** Existen diferentes implementaciones posibles de controles de calidad de datos dependiendo si los hechos reales que representan los datos son conocidos y accesibles o no. En caso que los hechos reales sean conocidos, los datos se pueden probar mediante validación (versus los registros fuente que contienen estos hechos), aunque la validación puede ser costosa y consumir mucho tiempo, y a menudo requiere el

uso de evaluadores si los hechos no están disponibles en forma estructurada y legible. Alternativamente, los datos se pueden probar mediante métricas de plausibilidad intrínsecas, evaluando el conjunto de datos con respecto a: otros datos del mismo conjunto de datos, diferentes rangos de referencia externos o otras tendencias plausibles (ciertos datos pueden ser válidos cuando se observan individualmente, pero la tendencia colectiva de todos los datos de un tipo debe seguir las distribuciones o tendencias esperadas).

## Conclusiones

La calidad de datos es un elemento crucial para realizar todo el potencial de la regulación basada en datos y respalda la confianza de los pacientes y profesionales de la salud. El proyecto de Marco de Calidad de Datos para la regulación de medicamentos en la Unión Europea, establece los criterios de calidad para los datos para garantizar que éstos sean aptos para el propósito de respaldar las decisiones de beneficio-riesgo.

Por otra parte, y con vista en el futuro, se prevé que el documento sea actualizado regularmente en concordancia con los nuevos desarrollos en la materia, incluyendo iniciativas para respaldar el Espacio Europeo de Datos de Salud (EHDS) propuesto por la Comisión Europea ●

## Normativas de referencia

- European Medicines Regulatory Network Data Standardisation Strategy. December 16th, 2021.
- Big Data Steering Group. Big Data Workplan 2022-2025.
- Good Practice Guide for the use of the Metadata Catalogue of Real-World Data Sources. EMA/787647/2022.
- Guideline on computerised systems and electronic data in clinical trials. EMA/INS/GCP/112288/2023.
- PIC/S Guidance. Good practices for data management and integrity in regulated GMP/GDP environments. PI 041-1. 1 July 2021.
- Good Practice Guide for the use of the Metadata. Catalogue of Real-World Data Sources. EMA/787647/2022. V 1.0 (draft)

## Acrónimos

DQF: Data Quality Framework  
 EMA: European Medicines Agency  
 HMA: Heads of Medicines Agencies  
 EU: European Union  
 MDM: Master Data Management  
 SGC: Sistema de Gestión de Calidad  
 ISO: International Organization for Standardization  
 EHDS: European Health Data Space