

Data Management Fundamentals

Módulo 10 – Data Quality



GOLD SPONSORS



SILVER SPONSORS



Entidades académicas





Lucía Engo Bermejo ✓

Head of Customer Success & Executive Committee | Estrategia de negocio y producto en Data & AI Governance



Universidad de Alcalá

www.linkedin.com/in/lucía-engo-bermejo



Data Management
Fundamentals



Data Quality



Índice de contenidos

1. Frameworks
2. Introducción a la calidad de datos
3. Conceptos esenciales
4. Actividades
5. Operaciones
6. Herramientas
7. Técnicas
8. Claves para la implementación
9. Calidad de datos y gobierno del dato
10. Resumen de puntos clave
11. Anexo

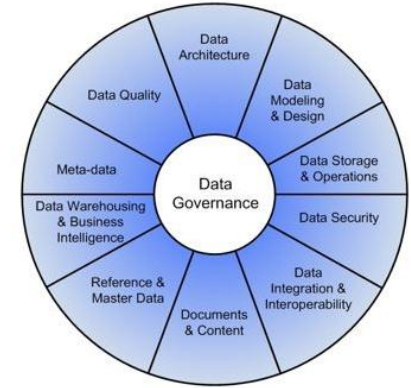


1. Frameworks

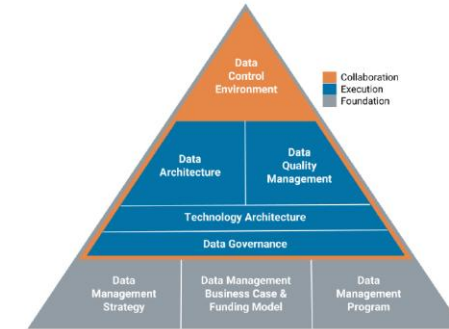
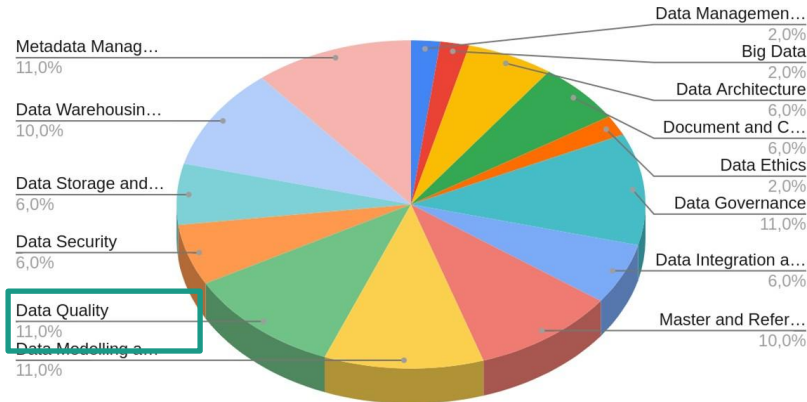
Frameworks

Principales frameworks con peso a nivel internacional:

- **DMBok2 (DAMA-I)**: www.dama.org
- **Data Management Capability Model –DCAM** – (Enterprise Data Management Council): <https://edmcouncil.org/>
- **Cobit (ISACA)**: <https://www.isaca.org/>
- **Data Governance Institute**: <http://www.datagovernance.com>



The DCAM™ Framework



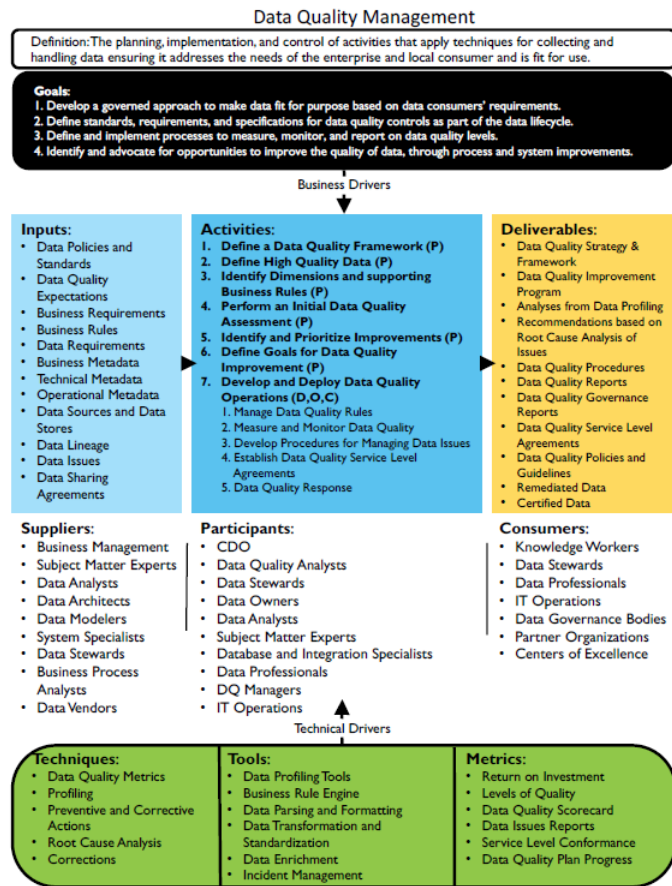
ISO/IEC 25012 Certificaciones de calidad de productos de datos



Especificación UNE 0079:2023 Gestión de la calidad del dato



Frameworks – DMBOK2R



(P) Planning, (C) Control, (D) Development, (O) Operations

Figura 91: Diagrama de Contexto de Calidad de Datos del DMBOK2R



La familia ISO 8.000 está compuesta por 4 partes principales:

- Conceptos generales sobre calidad de datos (partes 1, 2 y 8)
- Procesos de gestión de calidad de datos (partes 6x)
- Aspectos de intercambio de datos maestros entre organizaciones, atendiendo a ciertos aspectos de calidad (partes 100 a 150)
- Información de ingeniería (parte 311)

Este estándar incluye:

- Planificar qué es calidad de datos
- Controlar la calidad de datos
- Garantizar la calidad de los datos
- Mejorar la calidad de los datos



Mas información: <http://iso8000.es/normas-iso-8000>

2. Introducción a la Calidad de Datos

Introducción a la calidad de datos



El Data Management debe tener dos objetivos:

- Dar soluciones a las necesidades de negocio
- Contribuir positivamente a la calidad de datos

Aunque todas las disciplinas del Data Management contribuyen a la mejora de la calidad de los datos, es necesario un

Programa de Calidad de Datos:

- Gestión de la calidad a lo largo del ciclo de vida del dato
- Establecimiento de estándares
- Garantizar la calidad en los procesos que crean, transforman y almacenan los datos
- Contrastar la calidad frente a los estándares



El Programa de Calidad de Datos necesita:

- Equipo de calidad
- Participación en diferentes proyectos
- Interactuación con los diferentes stakeholders
- Identificación de best practices y mejoras en procesos
- Información y reportings sobre niveles de calidad
- Gestión de los metadatos de calidad
- Análisis, cuantificación, priorización y resolución de problemas

Recordatorio

La gestión de la calidad de los datos no es un proyecto con un inicio y un fin, es un **programa** que debe quedarse en la organización como cualquier otra función

El éxito de un programa requiere **comunicación, capacitación, compromiso, liderazgo**, así como un cambio de mentalidad donde la creación, el uso, el almacenamiento, procesamiento, etc tenga foco en la calidad de la información.



Los Business Drivers que llevan a una organización a emprender un programa de gestión de la calidad de los datos son:

- Mejorar la **experiencia** de los stakeholders y la **reputación** organizacional
- Mejorar la **efectividad** de la organización
- Reducir los **riesgos** y **costes** asociados a la mala calidad
- Mejorar la **eficiencia** y **productividad** de la organización

Para conseguir esto, se utiliza un enfoque basado en:

- Entender el contexto que impulsa la necesidad de los datos → **Fitness**
- Alinear la necesidad de calidad de los datos con este contexto → **Purpose**
- Asegurar que esta calidad es alcanzable mediante procesos estables



ROI de un programa de calidad:

Costes

- Personas
- Herramientas
- Datos externos
- Multas
- Denegación de servicio
- Sobre costes

Vs

Beneficios

- Mejora de la eficiencia de los procesos y proyectos
- Mejora de la reputación y/o incrementar el número de clientes
- Nuevos servicios y servicios basados en datos (Analytics, ML&AI)
- Venta de datos
- Toma de decisiones basadas en datos fiables

¿Cuánto le cuestan a las empresas sus problemas de calidad? ≡

Paquetes que no llegan porque falta la dirección, está incompleta o es errónea



Coste económico: Retrasos, reenvíos, logística ineficiente...



Satisfacción del cliente: Reclamaciones, quejas, pérdida de confianza



Impactos colaterales: Costes extras por almacenamiento, por deterioro de mercancía, retrabajo por gestión de devoluciones de pedidos



Los programas de calidad de datos se centran en conseguir los siguientes objetivos:



Enfoque de Gobierno

Conseguir que los datos cumplan con su propósito en base a las necesidades de los consumidores de datos



Estándares y Especificaciones

Definir estándares, especificaciones y requerimientos para establecer controles de calidad de los datos como parte del ciclo de vida de los datos



Medir, monitorizar e informar

Definir e implementar controles que permitan medir, monitorizar e informar sobre los niveles de calidad de los datos



Oportunidades de Mejora

Identificar oportunidades de mejora y abogar por la mejora continua de la calidad

Principios (1/2)



Los programas de calidad deben guiarse por los siguientes principios

Crítica

Enfoque en los datos más críticos para la organización, estableciendo prioridades basadas en valor y riesgo de los datos

Gestión del ciclo de vida

Gestión de la calidad en todo el ciclo de vida del dato, desde la creación, transformación, almacenamiento y eliminación

Prevención

Enfoque preventivo, centrado en evitar la aparición de errores, más allá de ceñirse a corregir datos

Remediación origen

Los problemas de calidad deben abordarse en sus causas origen en lugar de limitarse a remediar los síntomas. Para esto puede ser necesario realizar cambios en los procesos o los sistemas.

Gobierno

El gobierno debe respaldar la disposición de datos de calidad al igual que el programa de calidad debe respaldar y mantener un entorno de datos gobernado



Principios (2/2)



Los programas de calidad deben guiarse por los siguientes principios

Basado en
estándares

Todos los stakeholders tienen sus propios intereses y requerimientos a lo largo del ciclo de vida del dato, de forma que hay que establecer estándares y expectativas medibles contra los cuales se pueda medir la calidad

Medición
objetiva y
transparencia

Los niveles de calidad deben medirse de manera objetiva. Los resultados y la metodología aplicada deben compartirse con los diferentes stakeholders

Embebido en
procesos de
negocio

Los propietarios de los procesos de negocio son responsables de la calidad de los datos producidos a través sus procesos y deben hacer cumplir los estándares de calidad en sus procesos

Aplicación en
sistemas

Los propietarios de los sistemas deben hacer cumplir sistemáticamente los requerimientos de calidad de los datos

ANSs

En los Acuerdos de Nivel de Servicio se debe incorporar la monitorización y reporting así como la gestión de incidencias de calidad



A short white horizontal line in the upper left area of the slide.

3. Conceptos esenciales

“FITFORPURPOSE”

- Son adecuados para los fines a los que se desea aplicarlos
- Cumplen las expectativas de los usuarios consumidores
- Están disponibles cuando son requeridos

DATA QUALITY MANAGEMENT

“La planificación, implementación y control de actividades que aplican técnicas para recopilar y manejar datos, asegurando que aborden las necesidades de la empresa y del consumidor local y sean adecuados para su uso”

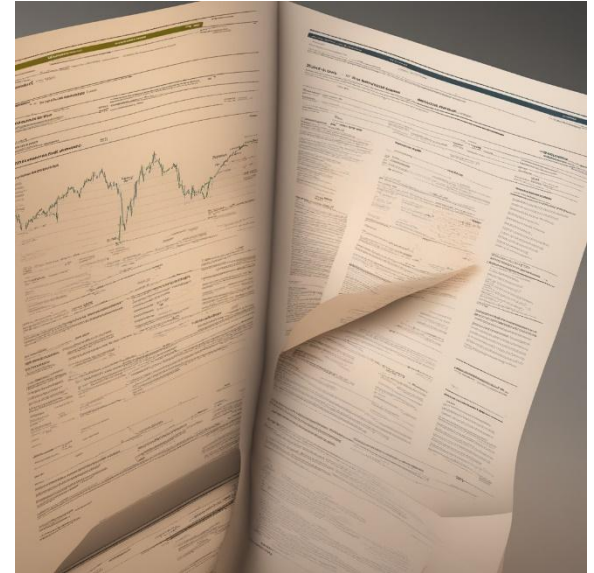


El concepto “Data Quality” engloba tanto las características asociadas con la alta calidad como los procesos utilizados para medir o mejorar la calidad de estos

El programa de calidad debe centrarse en los **datos críticos**, aquellos con mayor **relevancia** debido a importancia para negocio o por los **riesgos** que conlleva asociados.

Habitualmente los datos críticos están relacionados con:

- Informes regulatorios, financieros o de gestión
- Necesidades operativas de negocio
- Medidas de calidad del producto o satisfacción del cliente
- Estrategia para negocio, especialmente cuando se busca diferencia competitiva



Dimensiones de calidad



Una **dimensión de calidad** es una característica **medible** en los datos que proporciona un vocabulario común para definir y medir los requisitos de calidad de los mismos.

Las dimensiones de calidad estándar de DAMA-I se reflejan en la tabla:

Dimensión DQ	Descripción
Accuracy	La dimensión de exactitud refiere al modo en que el dato representa el mundo real
Completeness	La completitud refiere a la presencia de los datos requeridos
Consistency	La consistencia refiere a cómo un dato almacenado en un repositorio es consistente con otro dato almacenado en otro repositorio por estar relacionados
Integrity	La integridad de datos (o coherencia) incluye ideas asociadas a la completitud, exactitud y consistencia. Habitualmente se refiere a la integridad referencial (consistencia entre índices de tablas) o a la consistencia interna dentro de un conjunto de datos
Reasonability	Refiere a que el valor del dato está en línea con las expectativas y satisface los patrones esperados
Timeliness	La medida de disponibilidad refiere a la disponibilidad de un dato en el momento esperado o que se refresque con una determinada frecuencia para garantizar que el valor es vigente
Uniqueness/ Deduplication	La unicidad o des-duplicación establece que no existe más de una entidad en el mismo conjunto de datos
Validity	La validez indica si los valores son consistentes con los definidos dentro del dominio de valores definido



Dimensiones de calidad



MALA CALIDAD



- **Borroso o sin iluminación**
- **Sin metadatos**
- **No balanceado**

AI-READY

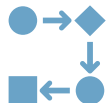


- **Nítido y bien iluminado**
- **Con metadatos** (fecha, cámara, ubicación, etiquetas...)
- **Balanceado y etiquetado**



Gestión de metadatos

Los metadatos son **críticos** para gestionar adecuadamente la calidad ya que contextualizan los datos (estándares, reglas, resultados de los controles, medidas de remediación, SLAs, tiempos de resolución...)



Requerimientos de datos

Es necesario disponer de un **proceso** sólido que permita documentar los **estándares** y requisitos de datos que permitan definir **reglas** con las que verificar la conformidad de forma periódica a través de una serie de **controles**.



Repositorio de metadatos de calidad

Los metadatos de calidad generados mediante los **controles** deben ser almacenados en un repositorio de metadatos y pueden ser integrados con otras herramientas de gobierno del dato (diccionario de datos, glosario de negocio...).

*** Información sobre seguimiento de incidencias y remediación*

Reglas de Negocio (Data quality business rules)



Las reglas de negocio describen cómo deben estar presentes los datos para servir al propósito de negocio.

Las reglas deben alinearse con las **dimensiones de calidad** y se utilizan para:

- Establecer validaciones de entrada
- Crear KQIs que permitan medir la calidad

Algunos ejemplos comunes de reglas:

- Conformidad de definición
- Presencia de valor y completitud de registro
- Cumplimiento de formato
- Pertenencia a un dominio de valores
- Conformidad de rango
- Conformidad de mapeo
- Verificación de precisión
- Verificación de unicidad
- Verificación de disponibilidad



Ciclo de vida de la mejora de la Calidad

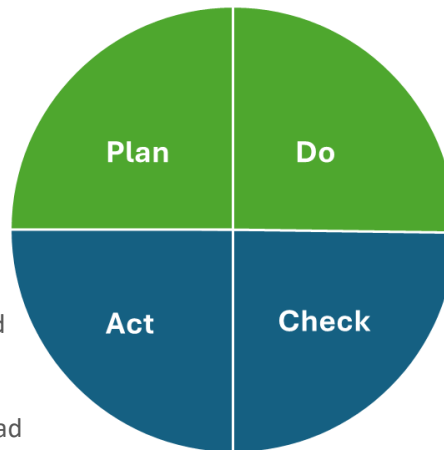


Enfoque basado en el modelo de resolución de problemas representado mediante el ciclo de Shewhart / Deming.

Inicio

Identificación de datos que no cumplen con los requisitos de los consumidores de datos o son un impedimento para la consecución de los objetivos de negocio

- Análisis de alcance, impacto y priorización de los problemas identificados.
- Evaluación de alternativas de remediación
- Es la etapa de actuar si:
 - $Calidad < \text{Umbral de Conformidad}$
 - Hay nuevos conjuntos bajo investigación
 - Surgen nuevos requisitos de calidad
 - Las reglas, estándares o expectativas cambian



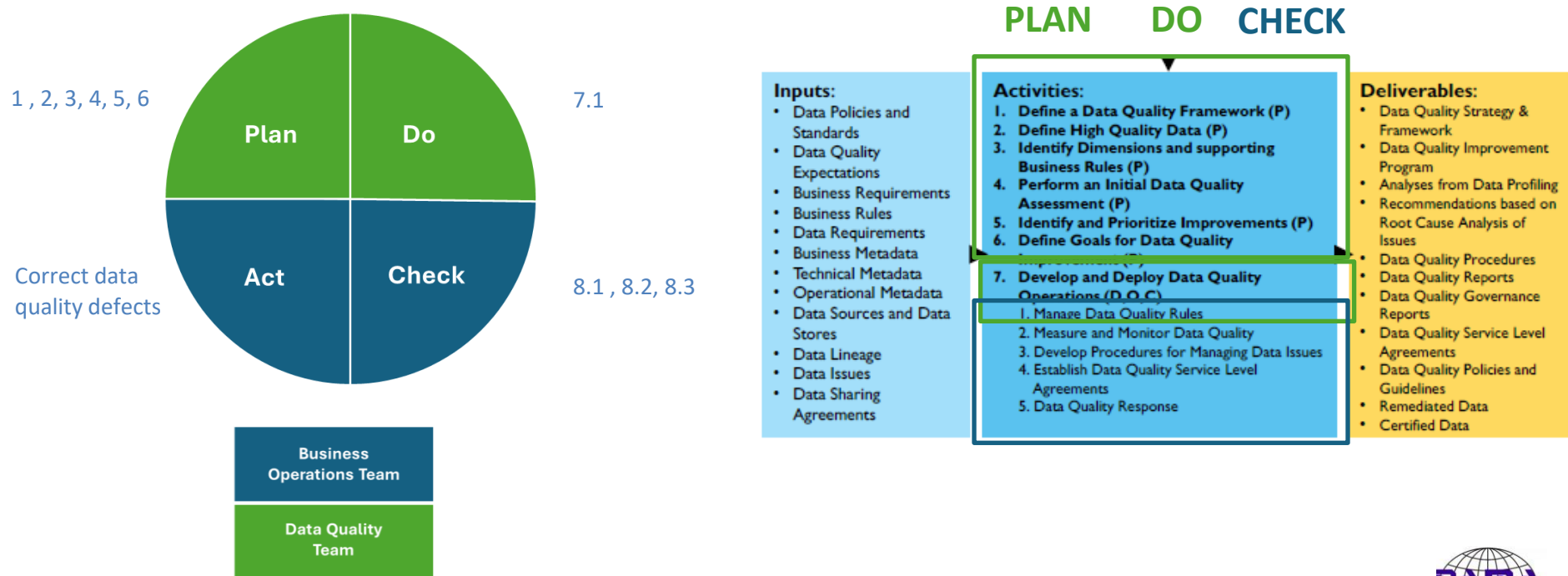
- Liderazgo de esfuerzos por parte del equipo de calidad para abordar las causas origen
- Planificación de la monitorización continua de la calidad
- Monitorización activa contra los requerimientos
- Verificación de los niveles de calidad conforme a los umbrales de aceptabilidad

Business
Operations Team

Data Quality
Team



Ciclo de vida de la mejora de la Calidad



Forma de análisis utilizada para inspección y descubrimiento de datos así como evaluación de la calidad.

Ejemplos de perfilado de datos:

- Cuentadenulos
- Valoresmáximosymínimos
- Longitud máximay mínima
- Distribución de frecuencia de valores para columnas individuales
- Tipo y formato de datos

El perfilado también incluye análisis entre columnas o tablas para:

- Identificar columnas duplicadas
- Identificar relaciones de claves foráneas (FK)

Mediante el análisis de los resultados de perfilado de datos por parte de un analista es posible:

- Descubrir datos
- Evaluar el estado actual de los datos
- Identificar problemas potenciales
- Analizar desviaciones
- Identificar oportunidades de mejora

Causas comunes de problemas de calidad



Las causas origen de los problemas suelen ser múltiples, y estas determinarán diferentes formas de prevenir problemas futuros.



Falta de Liderazgo

Los datos no se gestionan con riguro y hay disparidad de estructuras, formatos o valores que causan problemas de integración e interoperabilidad.



Procesos de Entrada

Falta de validaciones en interfaces de entradas que facilitan la entrada de errores al ecosistema informacional.



Funciones de Procesamiento

Suposiciones incorrectas sobre los orígenes de datos, reglas comerciales que están obsoletas, estructuras de datos modificadas.



Diseño del Sistema

Problemas al hacer cumplir la integridad referencial, genera duplicados, imprecisiones del modelo de datos, sobrecarga de campos, desajustes temporales, gestión débil de datos maestros, datos duplicados en silos.



Solucionar Problemas

Parches, comandos manuales o scripts creados rápidamente para arreglar datos incorrectos tienen un alto riesgo de causar errores aún mas graves.

Herramientas de soporte

- Catálogo de Datos
- Glosario de Negocio
- Flujos de aprobación
- Documentación actualizada sobre requerimientos
- Linaje y gestión de impactos
- Visión de arquitectura y monitorización de calidad
- Gestión del cambio

Calidad y Procesamiento de datos



Los esfuerzos de calidad deben ir enfocados en la prevención de errores, sin embargo, la calidad también puede mejorar a través del procesamiento de datos

Limpieza de Datos

- Se trata de detectar y corregir errores en los datos para alcanzar la calidad necesaria.
- Son costosas y no solucionan los problemas de origen
- Deberían reducirse a medida que se implementan controles, se corrigen los problemas en origen o se mejoran los procesos de negocio.

Enriquecimiento de Datos

- Proceso de agregar atributos a conjuntos de datos para aumentar su calidad y usabilidad. Se realiza mediante datos internos o adquiriendo datos externos a la organización. Ejemplos: fechas
- creación/modificación, información contextual, información geográfica ...

Análisis y formateo de Datos

- Proceso de revisar los datos para poder definir su contenido o valores y así establecer patrones contra los que determinar si un dato es válido o no y automatizar correcciones.
- Ejemplo: DNI con relleno delante y letra mayúscula, cuentas bancarias sin espacios...

Estandarización de Datos

- Proceso de transformación de datos basado en reglas establecidas a partir del conocimiento de los datos. Permite detectar patrones y transformar los datos a un estándar.
- Ejemplo: Alameda, avenida, calle, paseo...



A dark blue background with a faint, symmetrical pattern of light blue lines and circular shapes, resembling a stylized molecular structure or a network diagram.

4. Actividades

Actividades

1. Definir un framework de Calidad de Datos
2. Define “datos de alta calidad”
3. Identifica dimensiones de calidad y reglas de negocio
4. Realiza una evaluación inicial de la calidad → Piloto
5. Identifica y prioriza mejoras potenciales → Casos de uso
6. Define los objetivos de mejora de la calidad
7. Desarrolla las operaciones de mejora de la calidad de los datos



- El resultado del ASIS:
 - Visión del estado de la calidad en organización y principales problemas (painpoints, riesgos...)
 - Identificación de oportunidades: Business Drivers y áreas afines
 - Comprensión de la estrategia empresarial y los objetivos de negocio
 - Visión sobre el ecosistema informacional
 - Visión sobre la preparación cultural
 - Visión sobre qué mecanismos de gobierno hay en marcha para mejorar la calidad

- Para obtener una imagen completa del estado actual de la calidad de los datos es necesario:
 - Comprender la estrategia
 - Entrevistas con diferentes audiencias (Negocio e IT, diferentes niveles)
 - Documentación sobre dependencias de datos en los procesos de negocio
 - Documentación sobre la arquitectura técnica y soporte de sistemas a los procesos de negocio
 - Análisis de datos (perfilados y otras formas de análisis)

- Ejemplos:
 - ✓ ¿Qué quieren decir los stakeholders son “datos de calidad”?
 - ✓ ¿Cuál es el impacto de los datos de baja calidad?
 - ✓ ¿Qué prioridades impulsan la necesidad de mejorar la calidad?
 - ✓ ¿Qué estructura de gobierno existe para apoyar la mejora de la calidad?
 - ✓ ...



Definir un marco de referencia



La mejora de la calidad de los datos requiere una **estrategia** y adoptar un **marco de referencia** para conseguirlo. Un marco de referencia debe incluir métodos para:

GESTIONAR METADATOS

1. PRIORIZAR NECESIDADES

Comprender y priorizar las necesidades de negocio

2. IDENTIFICAR DATOS CRÍTICOS

Identificar los datos críticos para satisfacer las necesidades de negocio

3. DEFINIR REGLAS Y ESTÁNDARES

Establecer reglas de negocio y estándares de calidad de datos basados en los requisitos de negocio

4. EVALUAR DATOS

Evaluar los datos contra las expectativas

5. COMPARTIR RESULTADOS

Compartir los resultados y obtener feedback de los stakeholders para detectar problemas

6. PRIORIZAR PROBLEMAS

Priorizar los problemas detectados y gestionarlos adecuadamente

7. IDENTIFICAR OPORTUNIDADES

Identificar oportunidades de mejora a través del análisis de las causas origen, perfiles de datos, entrevistas a los stakeholders... Es importante definir **objetivos de mejora** en base a la criticidad, cantidad de datos afectados, riesgos, costes...

8. MONITORIZAR LA CALIDAD

Medir, monitorizar e informar sobre la calidad de los datos



Identificar Datos Críticos y definir Reglas

Los esfuerzos de gestión de la calidad de los datos deben centrarse en los datos más importantes de la organización

CRITERIOS PARA IDENTIFICACIÓN DE DATOS CRÍTICOS



DISTRIBUCIÓN



NATURALEZA



EXPLOTACIÓN



VALOR

1. PRIORIZAR NECESIDADES

2. IDENTIFICAR DATOS CRÍTICOS

3. DEFINIR REGLAS Y ESTÁNDARES

4. EVALUAR DATOS

5. COMPARTIR RESULTADOS

6. PRIORIZAR PROBLEMAS

7. IDENTIFICAR OPORTUNIDADES

8. MONITORIZAR LA CALIDAD

9. GESTIONAR METADATOS

Una vez identificados, se deben establecer **reglas de negocio** que describan las expectativas de calidad. Para definir reglas, es necesario:

- Comprender la finalidad para la que se crearon los datos
- Conocer los usos previstos para esos datos
- Relacionar los usos conocidos con las dimensiones de calidad (completitud, unicidad, validez...)

A partir de las **reglas de negocio** es posible establecer **reglas de calidad**:

- **A nivel de registro**: reglas de conformidad de mapeo, consistencia, completitud etc entre columnas de un mismo registro
- **A nivel de campo**: reglas generalmente sencillas basadas en completitud, validez en función de un rango de dominios, etc a nivel de columna
- **A nivel de conjunto**: reglas generalmente mas complejas porque relacionan diferentes tablas (p.e, que cada cliente tenga una dirección válida, el sumatorio de gastos no sea mayor que el presupuesto...)

Medir y Monitorizar la Calidad (1/2)

Los resultados de las reglas deben compararse contra un objetivo o umbral de calidad.

$$KQI \text{ válido}(r) = \frac{(Ejecuciones \text{ prueba } (r) - \text{Excepciones encontradas } (r))}{Ejecuciones \text{ prueba } (r)}$$

Donde (r) representa la regla

1. PRIORIZAR NECESIDADES
2. IDENTIFICAR DATOS CRÍTICOS
3. DEFINIR REGLAS Y ESTÁNDARES
4. EVALUAR DATOS
5. COMPARTIR RESULTADOS
6. PRIORIZAR PROBLEMAS
7. IDENTIFICAR OPORTUNIDADES
8. MONITORIZAR LA CALIDAD
9. GESTIONAR METADATOS

Dimensión	Regla de Negocio	Medida (Measure)	Métrica (Metric)	Indicador de Estado
Completitud	La presencia de un valor en un campo es obligatorio	Contar numero de registros completos y comparar con el total	$\frac{\# \text{ Registros rellenos}}{\# \text{ Total Registros}} \times 100$	Umbral de aceptabilidad: >80%
Completitud	El campo DNI es obligatorio	Campos con DNI: 790 Clientes totales: 1.000	$\frac{790}{1.000} \times 100 = 79\%$	Inaceptable
Unicidad	Debe existir un único registro por instancia de entidad en una tabla	Contar numero de registros duplicados y comparar con el total	$\frac{\# \text{ Registros duplicados}}{\# \text{ Total Registros}} \times 100$	Umbral de aceptabilidad: <0,01%
Unicidad	El cliente es único y se identifica por el DNI. El NIF del cliente solo puede aparecer una vez	Cantidad de NIF duplicados: 10 Clientes totales 1.000	$\frac{10}{1.000} \times 100 = 1\%$	Inaceptable

Medir y Monitorizar la Calidad (2/2)



Dimensión	Regla	Medida	Métrica	Indicador de Estado
Oportunidad	Los registros deben llegar en la ventana programada	Contar numero de registros que no llegan en ventana y comparar con el total	$\frac{\# \text{Registros fuera ventana}}{\# \text{Total Registros}} \times 100$	Umbral de aceptabilidad: <1%
Oportunidad	Datos covid de CCAA que llegan antes de las 10:00	Datos antes 10h : 7 Datos después: 5 Datos que no llegan: 3	$\frac{7}{17} \times 100 = 41,2\%$	Inaceptable
Validez	Si el campo X valor N, entonces campo Y valor N'	Contar numero de registros donde se cumple la relación	$\frac{\# \text{Registros cumplen}}{\# \text{Total Registros}} \times 100$	Umbral de aceptabilidad: >99%
Validez	Si provincia =Madrid, entonces CP=28***	Cantidad que cumple: 10k Cantidad que no cumple: 10	$\frac{10.000}{10.010} \times 100 = 99,9\%$	Aceptable

Reglas de mayor complejidad

- **Duplicados de cliente:** personas identificada con CNR que con el tiempo consiguen un NIF
- **Datos de fallecidos de covid:** Un día no se informan a tiempo 2, hoy se informa 1, ¿tabla de fallecidos acumulados+1 ó +3?
- **El prefijo del teléfono dependiente del país**
- **Dominios posibles de una cuenta de correo electrónico**



Gestionar Problemas de Datos (1/2)

1. PRIORIZAR NECESIDADES
2. IDENTIFICAR DATOS CRITICOS
3. DEFINIR REGLAS Y ESTÁNDARES
4. EVALUAR DATOS
5. COMPARTIR RESULTADOS
6. PRIORIZAR PROBLEMAS
7. IDENTIFICAR OPORTUNIDADES
8. MONITORIZAR LA CALIDAD
9. GESTIONAR METADATOS

Cuando los miembros del equipo de Calidad de Datos analicen los resultados, deben realizar las siguientes actividades

DIAGNÓSTICO DE PROBLEMAS

El objetivo es revisar los datos, rastrear el linaje, identificar el problema y las causas. Procedimiento:

- Revisar los datos en su contexto
- Evaluar si se han producido cambios que causen errores en el sistema
- Evaluar si existen otros problemas que contribuyan
- Determinar la posibilidad de causas externas

FORMULACIÓN DE OPCIONES DE REMEDIACIÓN

A partir del diagnóstico, evaluar distintas alternativas para abordar el problema. Estas pueden incluir:

- Abordar causas de raíz no técnica (falta de responsabilidades, liderazgo, capacitación)
- Modificaciones en sistemas
- Establecer controles para prevenir futuros problemas
- Corrección directa de datos erróneos
- No tomar medidas ya que el balance coste vs beneficio no compensa

RESOLUCIÓN DE PROBLEMAS

Una vez identificadas todas las posibilidades de remediación, consensuar con los stakeholders la mejor manera. Para ello:

- Evaluar costes relativos y ventajas de cada alternativa
- Recomendar una de las alternativas
- Proporcionar el plan de desarrollo e implementación
- Implementar la solución escogida



Identificar y Priorizar Mejoras

Una vez probado el éxito del proceso de mejora en un alcance acotado, el siguiente objetivo debe ser aplicarlo estratégicamente manteniendo la premisa de mejorar los datos críticos.



Mecanismos para identificar posibles mejoras

- **Perfilado de datos a gran escala** para comprender la amplitud de los problemas existentes
- **Entrevistas** a los stakeholder sobre los problemas de datos que les afectan
- **Seguimiento** de los problemas identificados y cómo estos impactan en negocio



Priorización de mejoras

La priorización de mejoras requiere una combinación del análisis de los datos y la discusión con las partes interesadas



Identificación de causas origen

- La identificación de problemas no es sinónimo de identificación de las causas raíz y tampoco permite determinar el impacto en los procesos de negocio.
- Determinar el impacto requiere la participación de todos los interesados a lo largo del ciclo de vida del dato

1. PRIORIZAR NECESIDADES
2. IDENTIFICAR DATOS CRÍTICOS
3. DEFINIR REGLAS Y ESTÁNDARES
4. EVALUAR DATOS
5. COMPARTIR RESULTADOS
6. PRIORIZAR PROBLEMAS
- 7. IDENTIFICAR OPORTUNIDADES**
8. MONITORIZAR LA CALIDAD
9. GESTIONAR METADATOS

Definir Objetivos para Mejorar la Calidad



El conocimiento obtenido será la base para establecer los objetivos de mejora a lograr mediante el Programa de Calidad.

Mecanismos de mejora



- Corrección de datos
- Remediación a corto-medio plazo
- Planes estratégicos

Obstáculos



Muchas cosas pueden obstaculizar los esfuerzos de mejora: limitaciones del sistema, resistencia cultural al cambio...

Objetivos específicos y alcanzables



Para evitar que las restricciones existentes detengan el programa, establecer objetivos realistas basados en el valor que proporcionan a negocio

Si logramos estimar el ROI, podremos justificar las acciones necesarias para alcanzar los objetivos.
El ROI se puede determinar en base a la **criticidad**, **cantidad** de datos, **procesos de negocio** afectados, **número o tipo de clientes** afectados, **riesgos** asociados, **costes** asociados a la remediación y de remediación rápida

5. Operaciones de Calidad de Datos

Gestionar Reglas de Calidad de Datos

Al igual que con otros tipos de metadatos, las reglas y estándares de calidad de datos deben gestionarse adecuadamente.

**Documentadas
consistentemente**



Existencia de plantillas y estándares para documentar reglas

**Definidas en términos de
dimensiones de Calidad**



Mejora el entendimiento sobre lo que se está midiendo y favorece el proceso de medición

**Vinculado al impacto
en negocio**



El objetivo es medir el impacto de la mala calidad en el negocio

**Respaldo por el
análisis de datos**



Al detectar un problema, el origen puede estar en los datos o en la definición de la propia regla

**Confirmados por
SMEs**



La definición adecuada de las reglas procede del conocimiento sobre los procesos de negocio

**Accesibles a los
consumidores de datos**



El acceso a las reglas mejora la comprensión de los datos y asegura que estas están correctamente definidas

Medir y monitorizar la calidad

De acuerdo con el principio de transparencia, la evolución de la calidad y del programa deben ser informados mediante:

Cuadros de Mando De
Calidad

Vista a alto nivel de métricas globales reportadas a diferentes niveles de la organización (ROI)

Tendencias de
Calidad

Tendencia de la evolución temporal de la calidad de los datos

Métricas de
ANSs

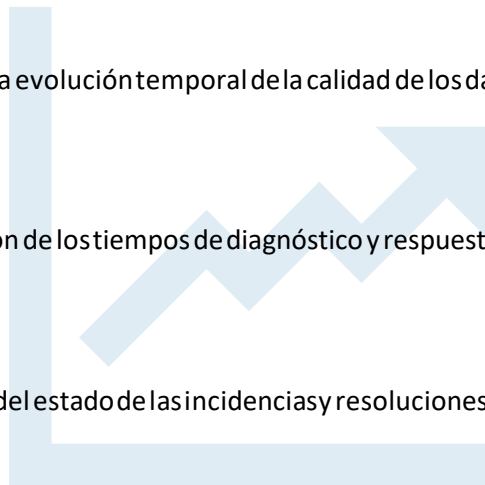
Representación de los tiempos de diagnóstico y respuesta a las incidencias

Gestión de
incidencias

Visualización del estado de las incidencias y resoluciones en curso

Efectos del
Programa

Efectos positivos de los proyectos de mejora del Programa de Calidad

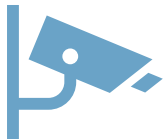


Procedimientos para gestión de incidencias



Seguimiento efectivo de incidencias

- 1 **Estandarizar** los problemas y las actividades de calidad
- 2 Proporcionar un proceso de **asignación** de problemas de datos
- 3 Administrar los procedimientos de **escalado** de problemas
- 4 Administrar el **flujo** de trabajo de **resolución** de incidencias de calidad



Las decisiones tomadas deben registrarse en un **sistema de seguimiento de incidencias**, que bien gestionado, puede proporcionar información útil sobre problemas, causas, costes, etc. Pero también proporcionará información sobre frecuencia, volumen de incidencias, responsables asociados, etc.

Los **consumidores** de datos también deben tener acceso a la información de seguimiento de incidencias para que así puedan tomar decisiones conociendo si los datos que utilizan han sido modificados y cómo han sido modificados.



Establecer Acuerdos de Nivel de Servicio



Los ANSs de Calidad especifican las expectativas de una organización para la respuesta y solución de los problemas de calidad de datos en cada sistema.

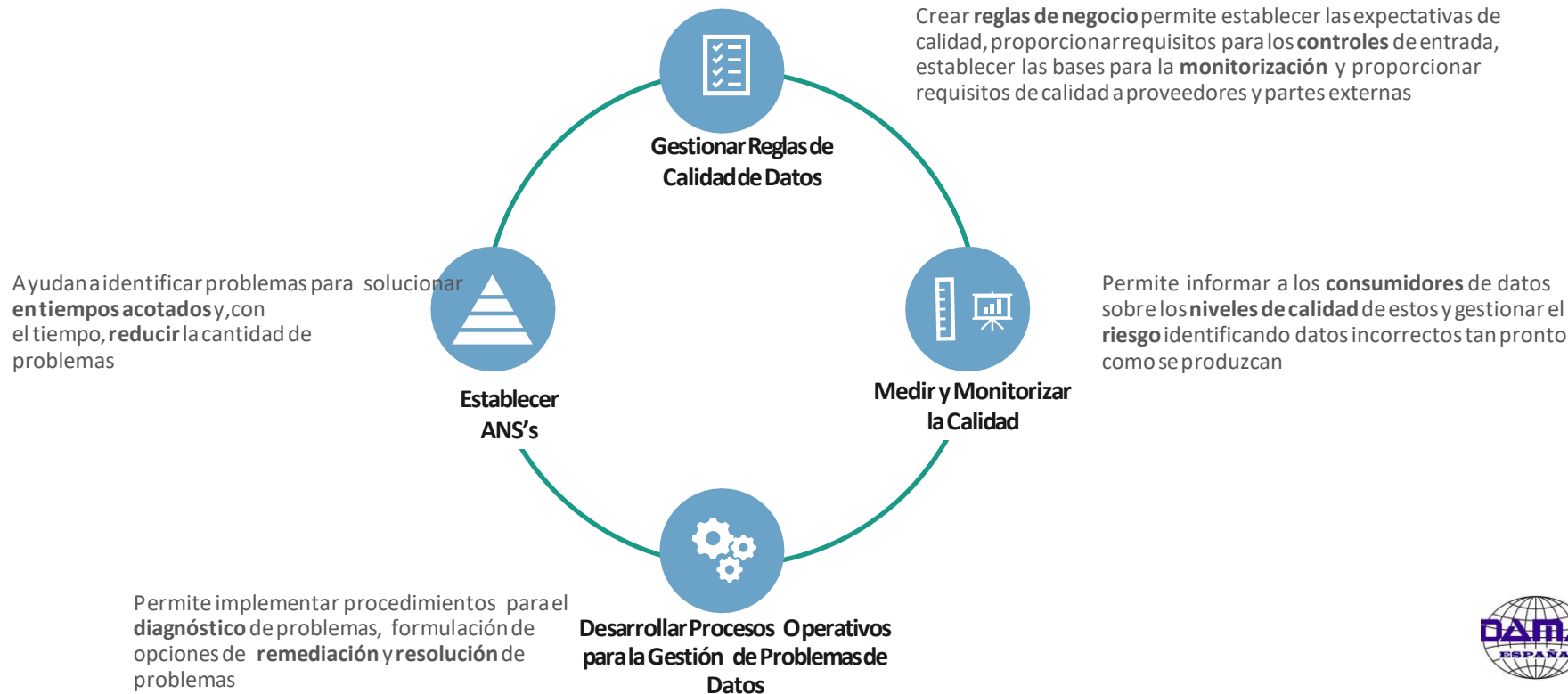
Los controles de calidad de datos operativos definidos en un SLA de calidad deben incluir:

- **Elementos de datos** cubiertos por el acuerdo
- **Impacto** en negocio asociado a los problemas de datos
- **Dimensiones** de datos asociadas a cada elemento de datos
- Expectativas de calidad para cada elemento de dato en cada una de las dimensiones identificadas en cada aplicación o sistema dentro de la cadena de valor de los datos
- Métodos para medir contra las expectativas
- Umbral de aceptabilidad para cada medición
- Responsable a ser notificado en caso de que no se cumpla el umbral de aceptabilidad
- Línea de tiempo y plazos de resolución
- Estrategia de escalamiento, posibles recompensas o penalizaciones

Desarrollar e Implementar Operaciones de Calidad



Muchos programas de calidad surgen de las mejoras identificadas a través de las actuaciones preliminares. Para mantener la calidad, el programa debe implementar un plan que desarrolle las siguientes actividades



A short, thick white horizontal line located in the upper left quadrant of the slide.

6. Herramientas

Herramientas de Perfilado de Datos



Dataformat profiling

Colu...	Data ...	Row ...	Null ...	Null %	Mean	Standard ...	Max
vitalperiodicid	int	100	0	0%	37407420.62	23854.9570363827	3
patientunitstayid	int	100	0	0%	141168.0	0.0	
observationoffset	int	100	0	0%	1230.5	636.2720893746	
temperature	double	100	100	100%	null	null	
sao2	int	100	73	73%	88.3703703704	11.0179858139	
heartrate	int	100	0	0%	108.38	22.6018592245	
respiration	int	100	100	100%	null	null	
cvp	int	100	94	94%	30.5	2.5099800796	
etco2	int	100	100	100%	null	null	
systemicsystolic	int	100	91	91%	109.2222222222	9.9219173774	

Rows per page: 10 1-10 of 19

DONE

CANCEL

Proporcionan estadísticas a alto nivel que permiten:

- Identificación de patrones en los datos
- Evaluación inicial de las características de calidad
- Autodescubrimiento de datos
- Monitorización continua de los datos



Plantillas de reglas de negocio y motor de ejecución

Ayudan a identificar posibles problemas mediante la ejecución de consultas para descubrir y cuantificar aspectos relacionados con las dimensiones de calidad.

 Create Rule

✓ Basic Rule Data

2 Specific Rule Data

3 Execution Conditions

☒ Null values are accepted as correct

☒ Value Range Selection

VALUE RANGE

Range Type

between

Min. Value Range

0

Max. Value Range

100

BACK

NEXT

CANCEL

Herramientas de Consulta de Datos



Las consultas pueden ser a nivel de columnas, registro o tablas.

Edit Rule

Threshold Configuration

Create New Join

SELECT A DATAFORMAT FROM THE REPOSITORY

sample-renfe-trips-500k.csv

Aggregate Dataset

Tag dataset

Make sure enrichment datasets do not contain duplicated values in the join columns or it may introduce duplicate rows in the final dataset

Unique Alias

Alias

Sales

Join Type

Select Join Type

left_join

Column Mapping 1

Select origin datasource column

id

Select Enriched datasource column

id

ADD NEW MAPPING

SAVE JOIN

CANCEL

Dataset Enrich

ADD JOIN

NEXT

CANCEL

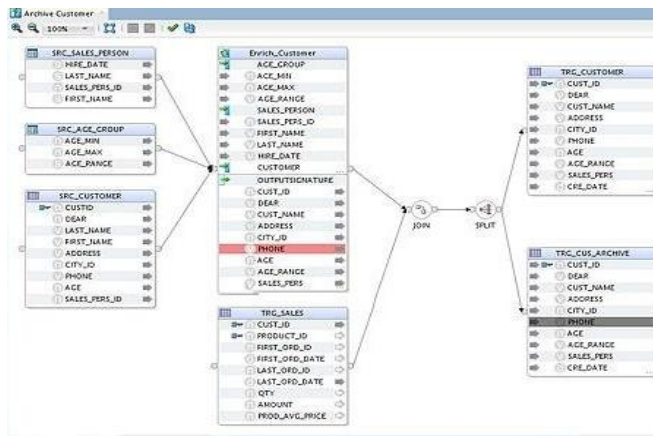


Herramientas Data Parsing y Formateo de datos

Existen diferentes herramientas para el *data parsing* (análisis y extracción de datos) y *data formatting* (formato y transformación de datos).

Las herramientas de **data parsing** permiten **analizar** el **contenido** de los datos para establecer **patrones** que después van a **motores** que permiten identificar **datos válidos** de los **no válidos**.

Cuando se identifican datos no válidos, se pueden **transformar** en datos válidos a partir de reglas basadas en errores conocidos.



Transformación y estandarización



TRANSFORMACIÓN Y ESTANDARIZACIÓN

Son herramientas que permiten **homogeneizar** datos conforme a **estándares** definidos para asegurar que los datos sean consistentes, interoperables y utilizables dentro y fuera de la organización.

ENRIQUECIMIENTO

Son herramientas que permiten **agregar** atributos a conjuntos de datos para aumentar su calidad y usabilidad.

GESTIÓN DE INCIDENCIAS

Son herramientas que permiten **registrar**, resolver y prevenir problemas en sistemas y servicios, mejorando la continuidad operativa.



A short white horizontal line.

7. Técnicas

A dark blue background with a faint, symmetrical pattern of light blue lines and circular shapes, resembling a stylized butterfly or a technical drawing.

Para gestionar adecuadamente la calidad de los datos, se dispone de diferentes técnicas:

Acciones que se implementan una vez se ha producido y detectado el problema. Las correcciones pueden ser automáticas, dirigidas manualmente o manuales.

Acciones para eliminar los factores que provocan la entrada de errores y así eliminar el problema en sí mismo.

Medibles, relevantes para negocio, comparables contra umbrales, deben tener un responsable, deben desencadenar acciones de remediación y deben medir la tendencia.



Definir métricas que informen a los consumidores sobre la calidad de los datos y las características relevantes de cara al consumo de esos datos.

Forma de análisis utilizada para inspección y descubrimiento de datos, así como evaluación de la calidad.

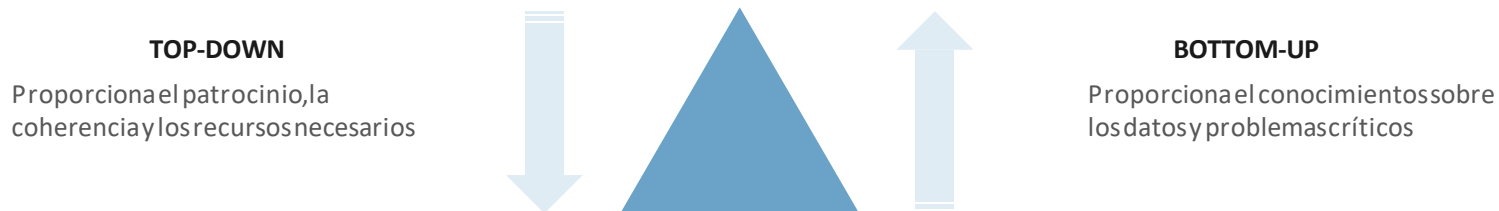
Acciones que evitan que ocurran errores conocidos.

8. Claves para la Implantación del Programa de Calidad

Claves para la Implantación



Gestionar la calidad de los datos no es algo sencillo, ni siquiera cuando se inicia desde un programa de GdD



Mejorar la calidad de los datos requiere un conjunto de cambios en la organización respecto a:

- El valor de los datos y el coste disponer de datos de baja calidad
- Modelo operativo y las interacciones entre TI y Negocio
- Modelo de ejecución de los proyectos
- Cambios en los procesos de negocio
- Presupuestos para remediación y mejora de la calidad
- Financiación para operaciones de monitorización continua de la calidad

Evaluación de la Preparación y Riesgos



El éxito de un programa de calidad viene determinado en gran medida por los apoyos de los que disponga. Para minimizar riesgos, es conveniente analizar la madurez en base a las siguientes características.



Compromiso de la organización para gestionar los datos como un activo estratégico



Conocimiento sobre los datos



Estado real de los datos



Conocimiento sobre los riesgos asociados a la creación, procesamiento y uso de los datos



Preparación cultural y técnica

9. Calidad y Gobierno de Datos

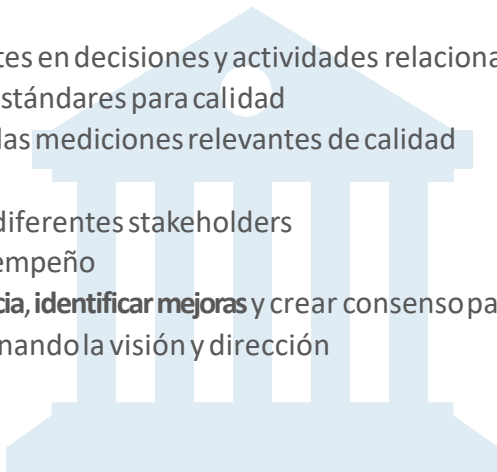
DQM y otras áreas del Dama Wheel



El programa de calidad es más efectivo cuando forma parte de un programa de gobierno del dato. Aunque en muchas ocasiones, los problemas de calidad son los catalizadores de la implantación de un modelo de gobierno.

El GdD puede acelerar el trabajo del programa de calidad mediante:

- Establecimiento de **prioridades**
- Identificación de **roles** participantes en decisiones y actividades relacionadas con la calidad
- Desarrollo y mantenimiento de estándares para calidad
- **Comunicaciones** relacionadas con las mediciones relevantes de calidad
- **Soporte** y orientación al personal
- **Formación** especializada para los diferentes stakeholders
- **Monitorización** e informes de desempeño
- Campañas para **promover conciencia, identificar mejoras** y crear consenso para las mejoras
- **Resolución de conflictos** proporcionando la visión y dirección



A dark blue background with a faint, abstract pattern of lines and circular shapes, resembling a technical drawing or a map.

10. Puntos clave

Lecciones aprendidas del módulo de calidad de datos:



Los datos de calidad son los que cumplen con los propósitos de negocio



La calidad requiere compromiso de todos los niveles de la organización Hay



que adoptar un enfoque preventivo y evitar la entrada de errores



Cuando se detecten errores, hay que actuar sobre las causas origen, no limitarse solo a corregir



Las reglas de calidad deben estar vinculadas a las reglas de negocio y asociadas a las Dimensiones de Calidad



La gestión de la calidad es un proceso cíclico en constante cambio

¡¡ Gracias !!



GOLD SPONSORS



SILVER SPONSORS



Entidades académicas



Universidad de
Castilla-La Mancha



MONDRAGON
UNIBERTSITATEA

