

Framework para la Gestión de la Calidad de Datos en Gobierno Digital

Proceso para la Gestión de Calidad de Datos

Temas

- ❑ Introducción
- ❑ Equipos y Roles
- ❑ Principales Etapas del Proceso
- ❑ Resumen Roles y Etapas

Los temas se presentan en forma de “guía rápida”. Por más detalles consultar el documento: “Framework para la Gestión de la Calidad de Datos en Gobierno Digital” [AGE2020].

Temas

- ❑ Introducción
- ❑ Equipos y Roles
- ❑ Principales Etapas del Proceso
- ❑ Resumen Roles y Etapas

Introducción



Marco Teórico



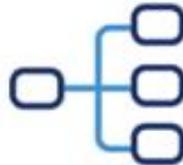
Proceso de Gestión de
Calidad de Datos



Marco Conceptual



Caso de estudio



Modelo de Calidad de
Datos de Referencia



Recursos de soporte



- ❑ El Proceso para la Gestión de Calidad de Datos del *framework* describe:
 - equipos y roles involucrados en la gestión de la calidad de datos en una organización y para escenarios específicos
 - principales etapas a seguir para gestionar la calidad de datos en un escenario específico

Temas

- Introducción
- Equipos y Roles
- Principales Etapas del Proceso
- Resumen Roles y Etapas



Equipos y Roles

Escenario

- ❑ Para aplicar el *framework* se sugiere la conformación de un **Comité de Calidad de Datos (CCD)** que será el responsable de llevar a cabo el proceso en un escenario



distintos roles pueden ser cubiertos por una misma persona

Equipos y Roles

Escenario



- ❑ Un CCD está conformado por roles en las áreas de calidad de datos, de negocio y técnicas [Wen07] [Int09]
- ❑ Estos roles y el CCD deben estar coordinados con otros equipos y roles vinculados a la gestión de datos en una organización, en particular con el Director de Datos de la Organización (Chief Data Officer, CDO) [Lee14][Bus16]

Equipos y Roles

Escenario



❑ Responsable de Calidad de Datos

- Responsable de la aplicación del *framework* en un escenario
- Está a cargo de todas las etapas del proceso y de dirigir las actividades de los otros integrantes del comité

❑ Analista de Calidad de Datos

- Experto en el área de calidad de datos
- Responsable de los aspectos de calidad de datos en escenario
 - ❖ p. ej. analizar requerimientos de calidad, identificar actores relevantes, identificar problemas de calidad, definir métricas de calidad

Equipos y Roles

Escenario



❑ Técnico de Calidad de Datos

- Experto en el área de calidad de datos con perfil técnico
- Responsable de recursos técnicos para aplicación del *framework*
 - ❖ p. ej. examinar datos objetivo, implementar métodos de medición

❑ Experto de Negocio

- Referente del área de negocio en la que se enmarca el escenario

❑ Experto Técnico

- Referente de aspectos técnicos vinculados al escenario

Equipos y Roles

Organización



❑ Director de Calidad de Datos de la Organización

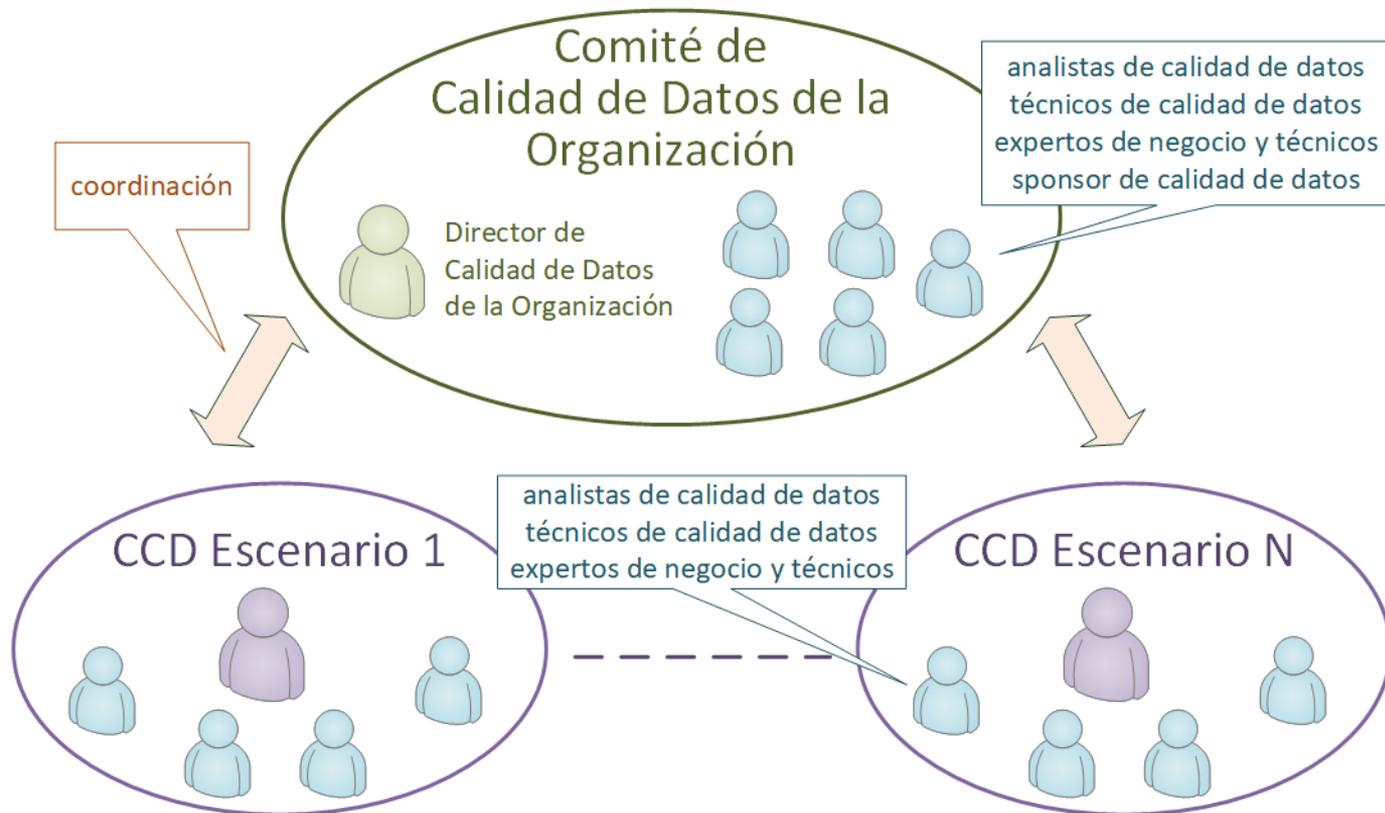
- Responsable de Calidad de Datos en la Organización
- Encargado de designar un CCD para cada escenario (junto con el CCD de la Organización)

❑ Comité de Calidad de Datos de la Organización

- Liderado por el Director de Calidad de Datos
- Integrado por expertos en calidad de datos, de negocio y técnicos
- Puede integrarse a este comité un Sponsor de Calidad de Datos
 - ❖ dirigir, financiar y supervisar la gestión de la calidad de datos

Equipos y Roles

Resumen



Temas

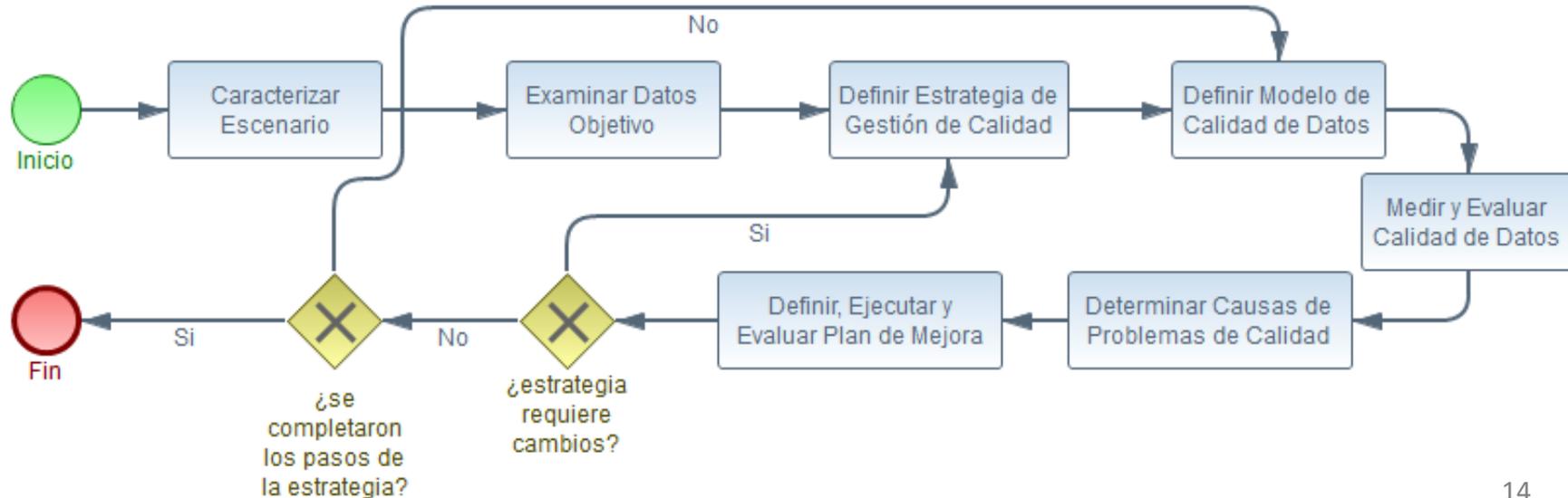
- Introducción
- Equipos y Roles
- Principales Etapas del Proceso
- Resumen Roles y Etapas



Principales Etapas del Proceso

Introducción

- El proceso consiste de siete etapas a ser llevadas a cabo por los distintos roles

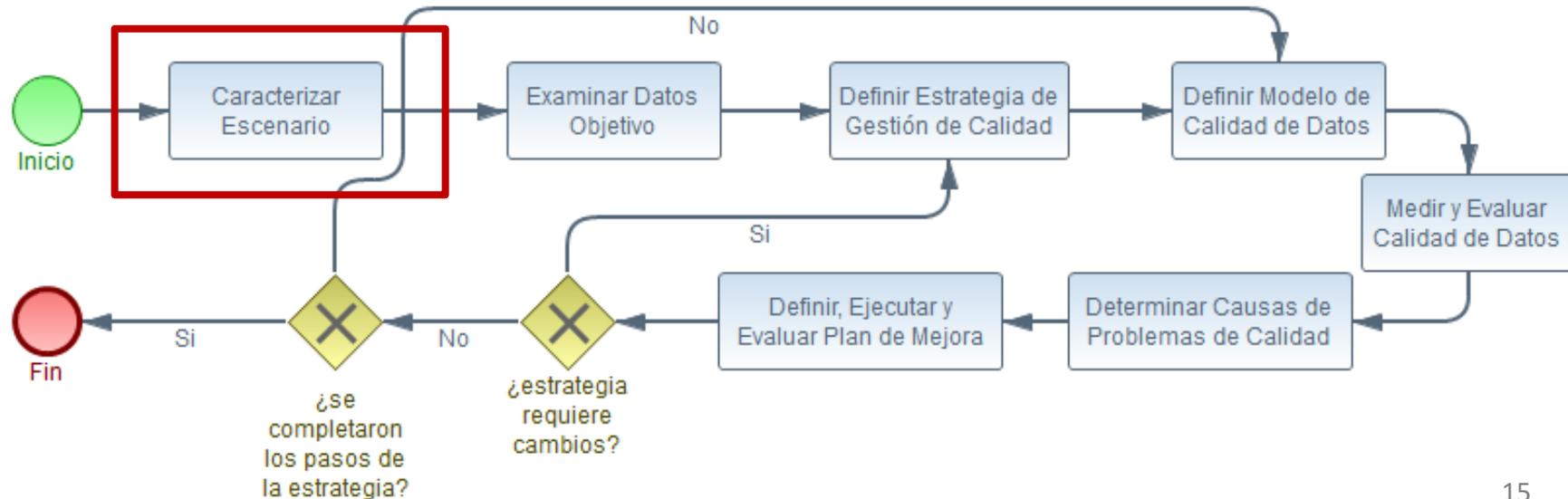




Principales Etapas del Proceso

Caracterizar Escenario

- El proceso consiste de siete etapas a ser llevadas a cabo por los distintos roles





Principales Etapas del Proceso

1: Caracterizar Escenario

- ❑ Tiene como objetivo identificar los elementos relevantes para el escenario de trabajo, desde dos puntos de vista:
 - Caracterización Técnica y de Negocio
 - ❖ p. ej. sistemas que procesan los datos, procesos de negocio
 - Caracterización de Calidad de Datos
 - ❖ p. ej. requerimientos y problemas de calidad de datos

- ❑ Estos elementos brindan insumos para las siguientes etapas del proceso

Principales Etapas del Proceso



1: Caracterizar Escenario: Aspectos Técnicos y de Negocio

- ❑ El objetivo es identificar los elementos técnicos y de negocio relevantes para el escenario
- ❑ Se apunta a determinar:
 - cómo los distintos artefactos tecnológicos (p. ej. aplicaciones) brindan soporte a las funcionalidades de negocio del escenario
 - el interés de distintos actores en los elementos del escenario

Principales Etapas del Proceso



1: Caracterizar Escenario: Aspectos Técnicos y de Negocio

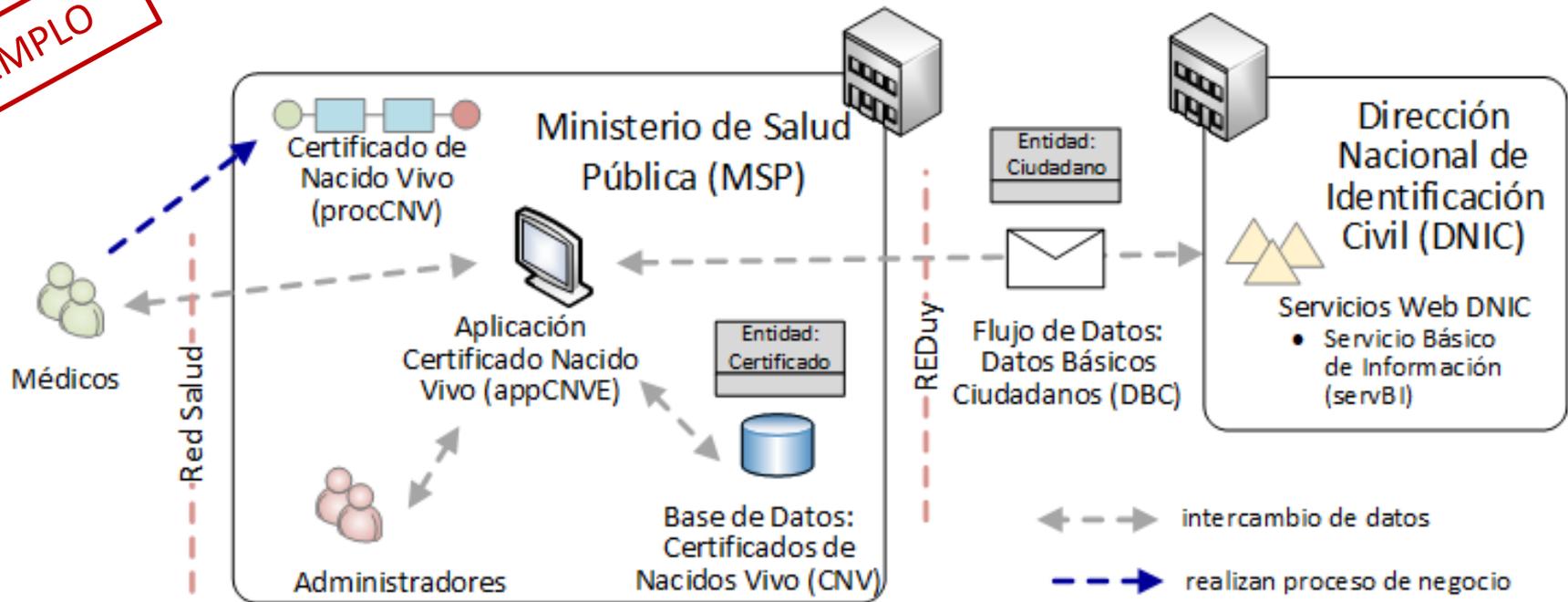
- ❑ Algunos de los elementos técnicos a identificar son:
 - aplicaciones que procesan datos
 - organizaciones con las que se intercambian datos
 - roles de usuario que utilizan las aplicaciones
 - colecciones de datos (p. ej. bases de datos) que utilizan estas aplicaciones
- ❑ Algunos de los elementos vinculados al negocio a identificar son:
 - los procesos y entidades de negocio relevantes para el escenario
 - las formas en que los elementos técnicos dan soporte a estos procesos y gestionan estas entidades
 - los actores relevantes y su vinculación con los distintos elementos identificados



Principales Etapas del Proceso

1: Caracterizar Escenario: Aspectos Técnicos y de Negocio

EJEMPLO

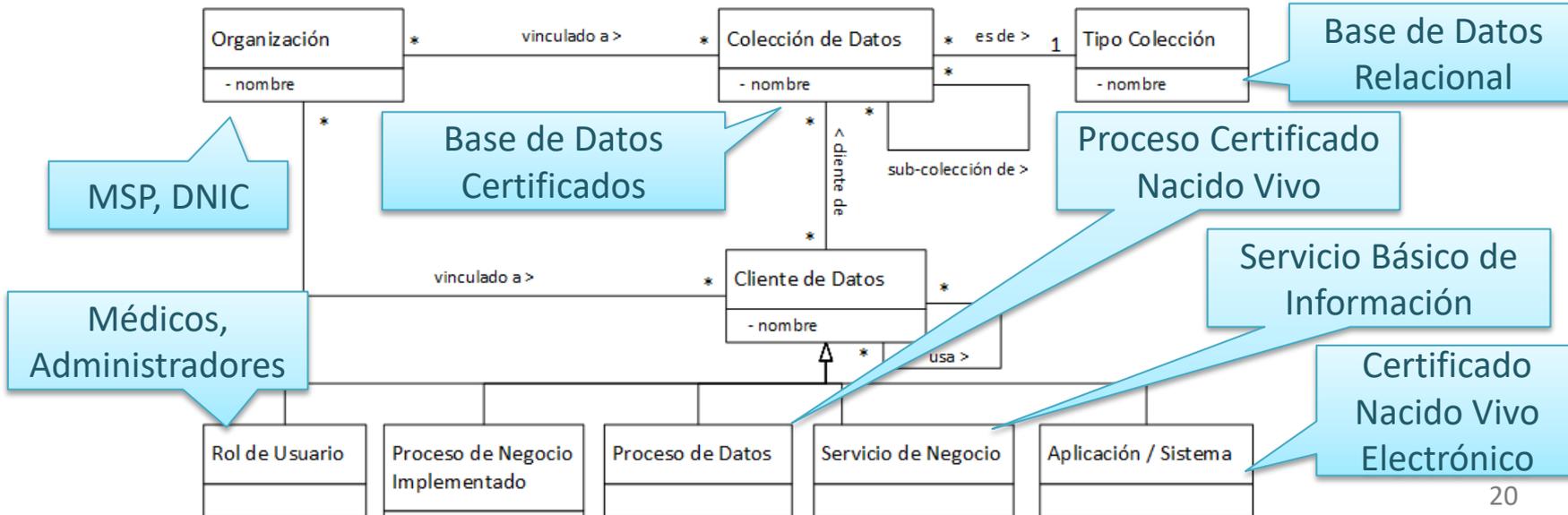




Principales Etapas del Proceso

1: Caracterizar Escenario: Aspectos Técnicos y de Negocio

- El Marco Conceptual del *framework* brinda una forma de especificar los elementos del escenario y sus relaciones



Principales Etapas del Proceso



1: Caracterizar Escenario: Aspectos Técnicos y de Negocio

- ❑ La caracterización técnica y de negocio permite:
 - tener una visión general de los elementos técnicos y de negocio
 - conocer la forma en que los elementos técnicos brindan soporte a las funcionalidades de negocio

- ❑ Esta caracterización es un insumo importante para etapas posteriores del proceso, por ejemplo:
 - para determinar causas de problemas de calidad
 - para priorizar elementos en la gestión de la calidad de datos

Principales Etapas del Proceso



1: Caracterizar Escenario: Aspectos Técnicos y de Negocio

- ❑ Los principales roles del CCD involucrados en esta etapa de caracterización técnica y de negocio son:
 - Experto de Negocio
 - Experto Técnico



Principales Etapas del Proceso

1: Caracterizar Escenario: Aspectos de Calidad de Datos

- ❑ El objetivo es identificar los elementos relevantes para el escenario desde un punto de vista de la calidad de datos
- ❑ Como resultado de esta etapa se espera identificar:
 - los requerimientos de calidad de datos
 - las relaciones entre requerimientos y elementos del escenario
 - los problemas de calidad de datos



Principales Etapas del Proceso

1: Caracterizar Escenario: Aspectos de Calidad de Datos

EJEMPLO



Requerimiento 1. La fecha de nacimiento en un certificado no puede ser posterior a la actual.



Requerimiento 2. El número de serie de un certificado no puede ser nulo.



Problema 1. Un gran número de certificados no tienen datos completos de la madre.



Problema 2. Los países se especifican de distinta forma en certificados (p. ej. Uruguay, URU y UY).



Requerimiento 3. Los nombres, apellidos y CI de la madre en un certificado deben estar especificados.

Actores



Gerentes
MSP



Salud.uy



Médicos

interés

detecta

interés

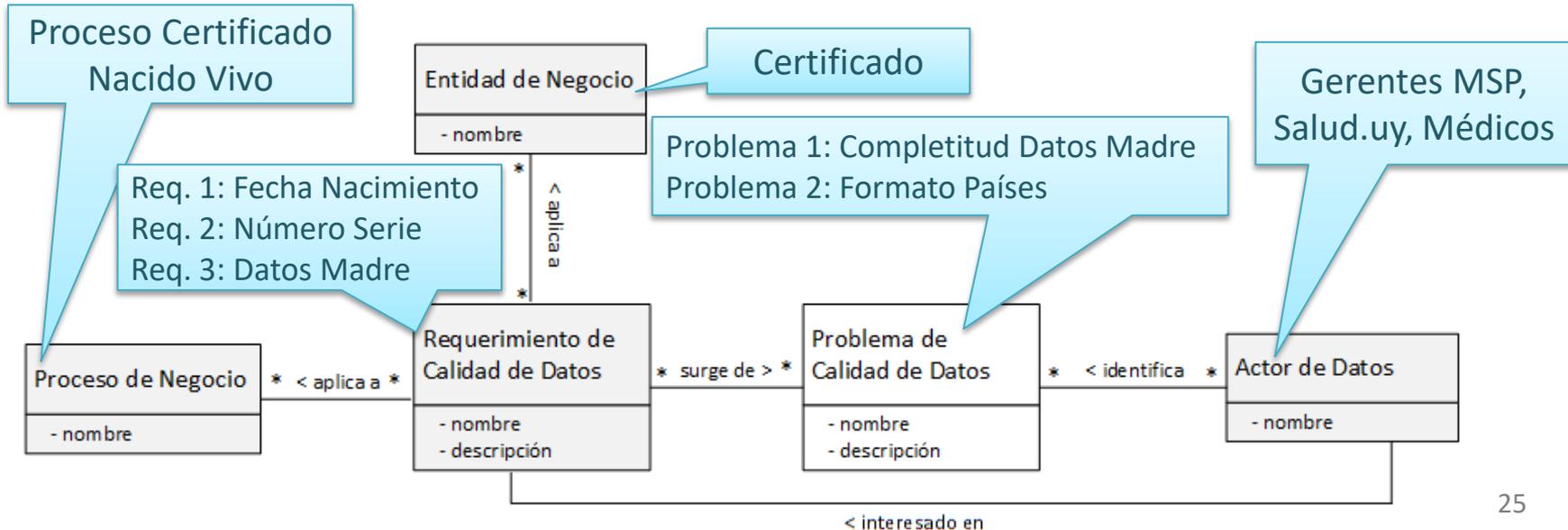
genera



Principales Etapas del Proceso

1: Caracterizar Escenario: Aspectos de Calidad de Datos

- El Marco Conceptual del *framework* brinda una forma de especificar los aspectos de calidad de datos del escenario.



Principales Etapas del Proceso



1: Caracterizar Escenario: Aspectos de Calidad de Datos

- ❑ Esta caracterización constituye un insumo importante para etapas posteriores del proceso, en particular, para definir:
 - la estrategia de gestión de calidad de datos
 - el plan de mejora

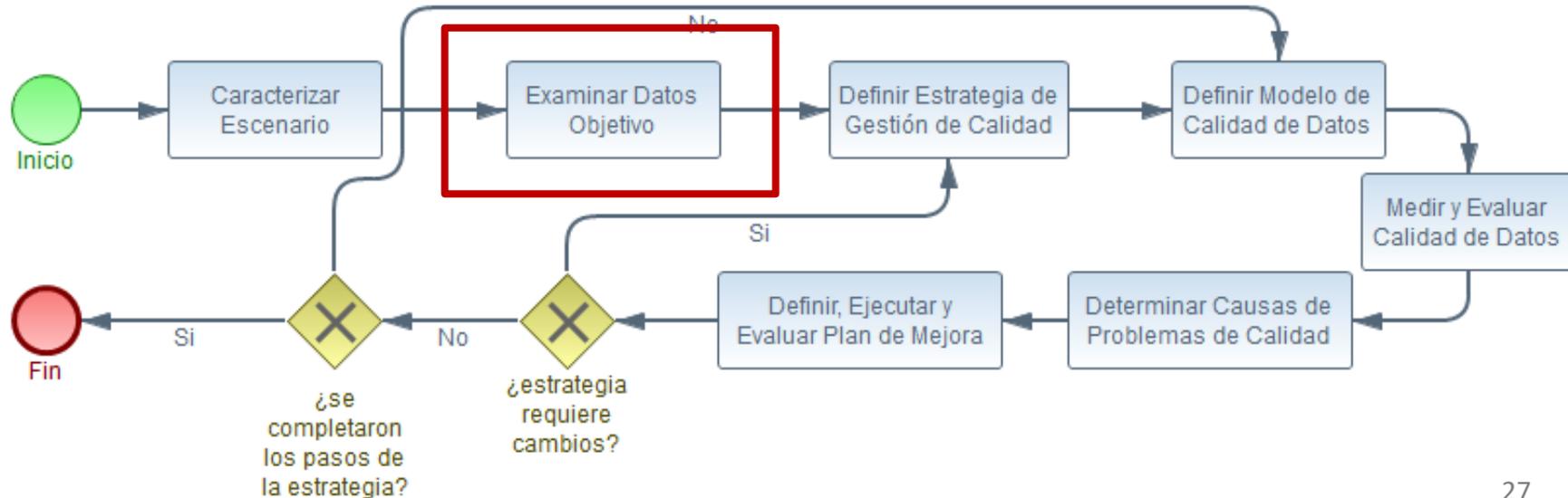
- ❑ Los principales roles del CCD involucrados en esta etapa de caracterización de calidad de datos son:
 - Experto de Negocio
 - Analista de Calidad de Datos



Principales Etapas del Proceso

2. Examinar Datos Objetivo

- El proceso consiste de siete etapas a ser llevadas a cabo por los distintos roles





Principales Etapas del Proceso

2. Examinar Datos Objetivo

- ❑ El objetivo es:
 - conocer las características de los datos
 - realizar una primera estimación de la calidad de estos datos
 - detectar problemas de calidad de datos
 - definir nuevos requerimientos a partir de estos problemas

- ❑ Una de las principales herramientas de soporte en esta etapa es la aplicación de técnicas de *Data Profiling*

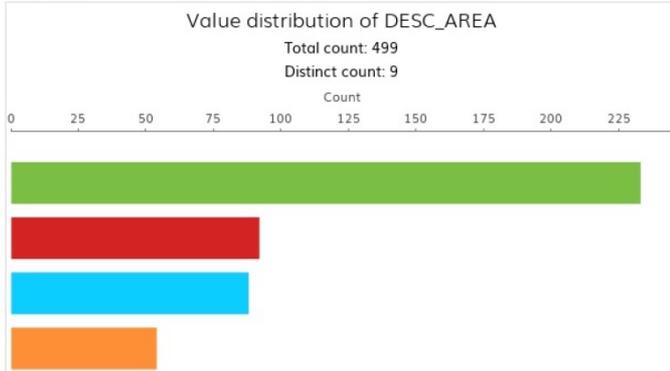


Principales Etapas del Proceso

2. Examinar Datos Objetivo: Técnicas Data Profiling

EJEMPLO

Value distribution
(DESC_AREA)



Value	CO...
Limpieza	233
Alumbrado	92
Sanearie...	88
Arbolado	54
CECOED	17
Calles y ve...	11
Espacios ...	2
Arbolado1	1
Barometri...	1

Number analyzer

Row count	498
Null count	0
Highest value	990
Lowest value	-1.574
Sum	25.993
Mean	52,195
Geometric mean	
Standard deviation	143,266
Variance	20.525,256
Second moment	10.201.052,106
Sum of squares	11.557.751

Pattern finder

(EMAIL)

	Match count	Sample
aaaaaaaaa@aaaaaaaaaaaaaaaaaaa.aaa	23 23	dmurphy@classic.com:



Principales Etapas del Proceso

2. Examinar Datos Objetivo

- ❑ Como resultado de esta etapa se apunta a conocer:
 - los tipos de las colecciones de datos y de los atributos de entidades
 - una primera estimación del alcance de los problemas y del cumplimiento de requerimientos
 - nuevos problemas y requerimientos de calidad detectados con técnicas de Data Profiling



Principales Etapas del Proceso

2. Examinar Datos Objetivo

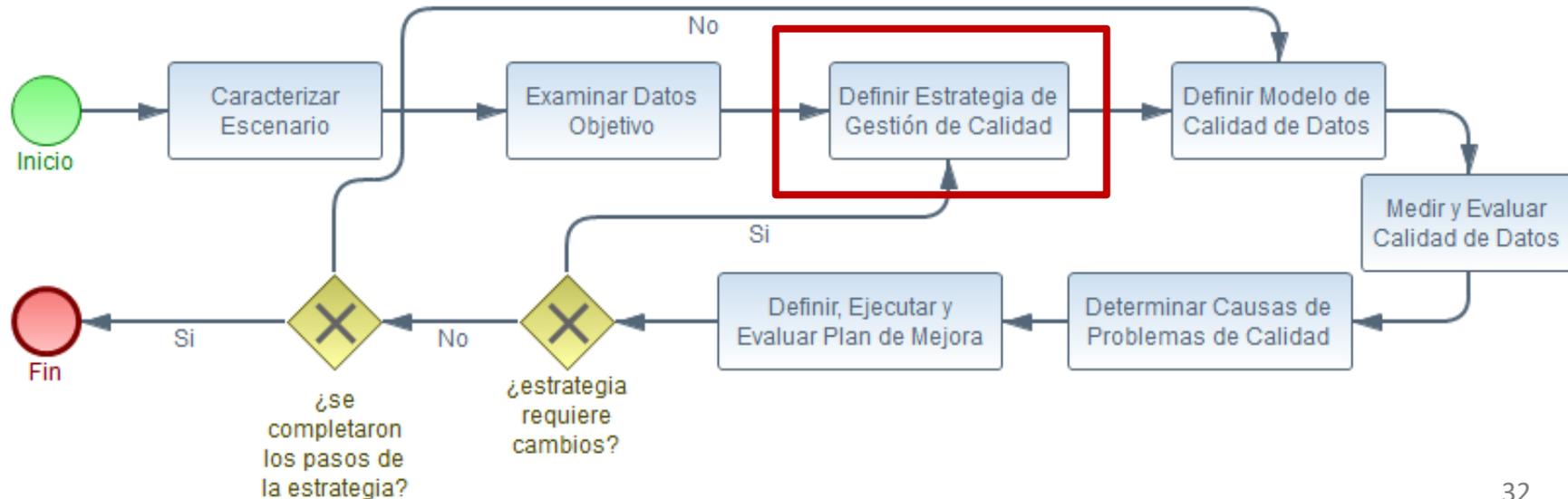
- ❑ Los principales roles del CCD involucrados en esta etapa de Examinar Datos Objetivo son:
 - Experto Técnico
 - Experto de Negocio
 - Técnico de Calidad de Datos



Principales Etapas del Proceso

3. Definir Estrategia de Gestión de Calidad

- El proceso consiste de siete etapas a ser llevadas a cabo por los distintos roles





Principales Etapas del Proceso

3. Definir Estrategia de Gestión de Calidad

- ❑ El objetivo es definir la estrategia para gestionar la calidad de datos en el escenario de trabajo
- ❑ Los resultados esperados de esta etapa son:
 - la priorización de los requerimientos
 - la definición de la estrategia en base a pasos que abordan un conjunto de requerimientos de datos



Principales Etapas del Proceso

3. Definir Estrategia de Gestión de Calidad

EJEMPLO

Priorización

MEDIA



Requerimiento 1. La fecha de nacimiento en un certificado no puede ser posterior a la actual.

ALTA



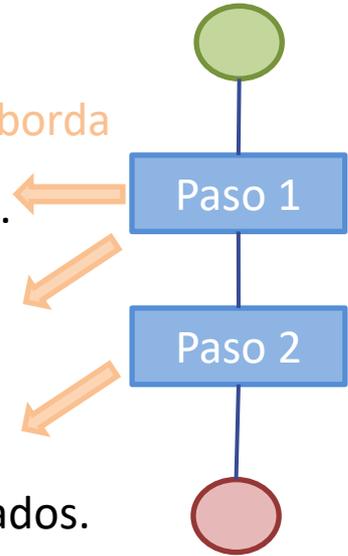
Requerimiento 2. El número de serie de un certificado no puede ser nulo.

BAJA



Requerimiento 3. Los nombres, apellidos y CI de la madre en un certificado deben estar especificados.

aborda



Definición Estrategia



Principales Etapas del Proceso

3. Definir Estrategia de Gestión de Calidad

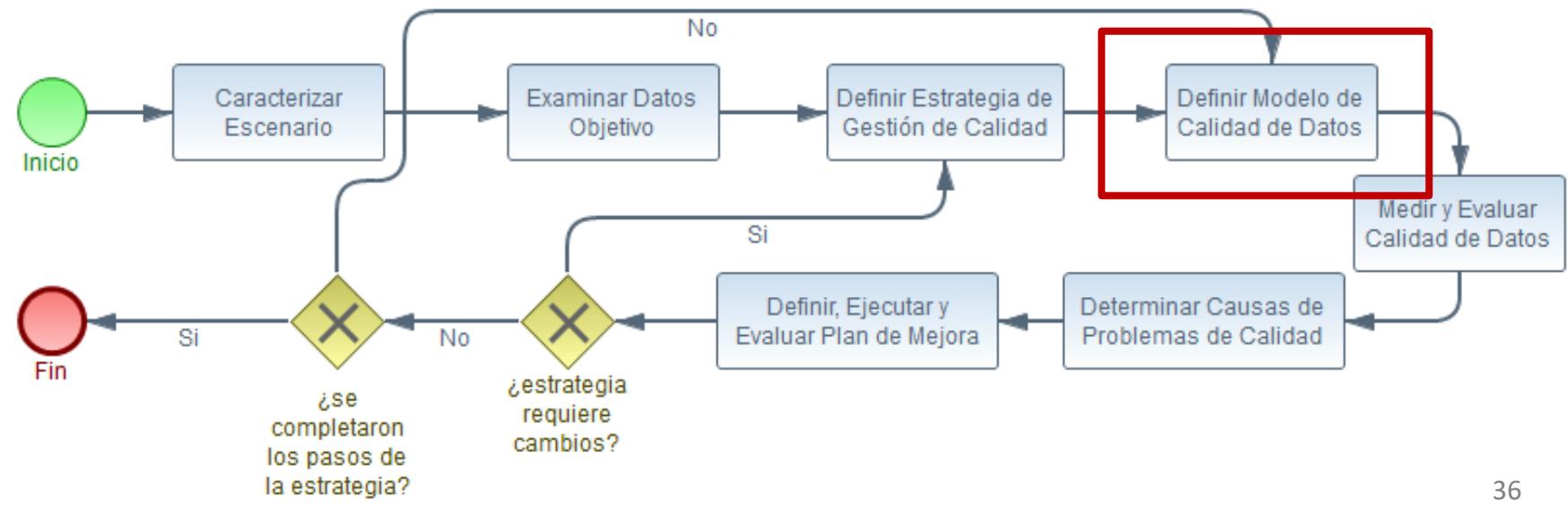
- ❑ Los principales roles del CCD involucrados en esta etapa de Definir Estrategia de Gestión de Calidad son:
 - Responsable de Calidad de Datos
 - Experto de Negocio
 - Analista de Calidad de Datos



Principales Etapas del Proceso

4. Definir Modelo de Calidad

- El proceso consiste de siete etapas a ser llevadas a cabo por los distintos roles





Principales Etapas del Proceso

4. Definir Modelo de Calidad

- ❑ El objetivo es definir el modelo de calidad de datos para el escenario de trabajo
- ❑ El modelo se construye en base a:
 - el modelo de calidad de datos de referencia
 - los requerimientos de calidad de datos considerados en la estrategia



Principales Etapas del Proceso

4. Definir Modelo de Calidad

- El resultado esperado de esta etapa es el modelo de calidad de datos para un escenario específico incluyendo:
 - la identificación de los elementos del modelo de referencia a utilizar
 - la definición de los elementos base del modelo para el escenario
 - ❖ dimensiones, factores, métricas
 - la definición de perfiles de evaluación



Principales Etapas del Proceso

4. Definir Modelo de Calidad



- Dimensiones
- Factores
- Métricas
- Métodos
- Reglas de Evaluación
- Perfiles de Evaluación



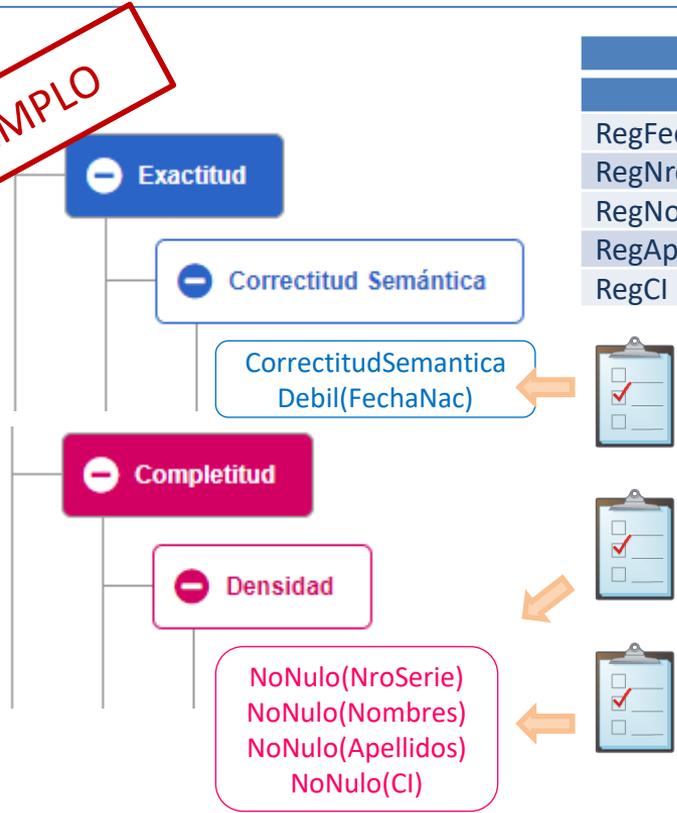
**Requerimientos de
Calidad de Datos
de Escenario**



Principales Etapas del Proceso

4. Definir Modelo de Calidad

EJEMPLO



Perfil Evaluación		
Regla	Métrica	Condición
RegFecha	CorrectitudSemanticaDebil(FechaNac)	resultado = true
RegNroSerie	NoNulo(NroSerie)	resultado = true
RegNombres	NoNulo(Nombres)	resultado = true
RegApellidos	NoNulo(Apellidos)	resultado = true
RegCI	NoNulo(CI)	resultado = true



Requerimiento 1. La fecha de nacimiento en un certificado no puede ser posterior a la actual.



Requerimiento 2. El número de serie de un certificado no puede ser nulo.



Requerimiento 3. Los nombres, apellidos y CI de la madre en un certificado deben estar especificados.



Principales Etapas del Proceso

4. Definir Modelo de Calidad

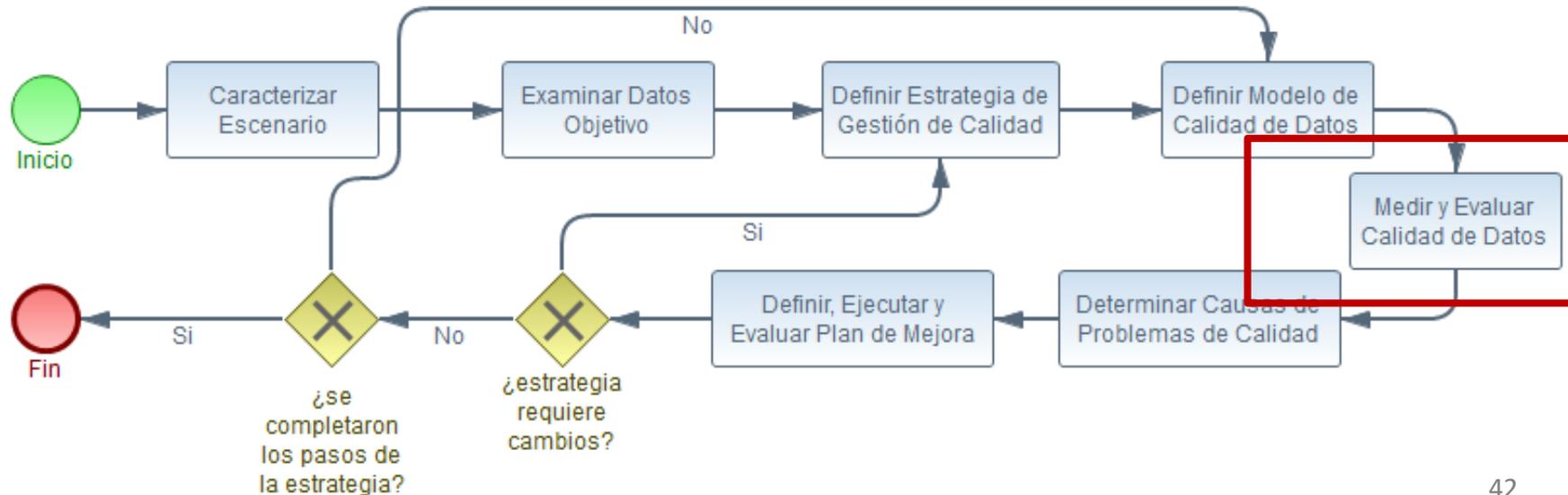
- ❑ Los principales roles del CCD involucrados en esta etapa de Definir Modelo de Calidad son:
 - Responsable de Calidad de Datos
 - Experto de Negocio
 - Analista de Calidad de Datos



Principales Etapas del Proceso

5. Medir y Evaluar Calidad de Datos

- El proceso consiste de siete etapas a ser llevadas a cabo por los distintos roles





Principales Etapas del Proceso

5. Medir y Evaluar Calidad de Datos

- ❑ El objetivo es:
 - realizar mediciones utilizando las métricas y métodos del modelo de calidad
 - evaluar la calidad de los datos en base a los perfiles y reglas de evaluación

- ❑ Los resultados esperados de esta etapa son:
 - implementación de métodos de medición para métricas que lo requieran
 - medidas para las métricas definidas en el modelo
 - evaluación de la calidad en base a medidas, perfiles y reglas de evaluación



Principales Etapas del Proceso

5. Medir y Evaluar Calidad de Datos

EJEMPLO

Medidas Métrica NoNulo(NroSerie)			
IdMedida	Fecha	Resultado	IdObjeto (Certificado)
1	04/05/2020	true	A123
2	04/05/2020	false	A124
3	04/05/2020	false	A125
4	04/05/2020	true	A126
5	04/05/2020	true	A127

El perfil de evaluación se cumple en un 68%, por lo que se deben determinar las causas para poder definir acciones.

Evaluación Perfil

0,68

evaluación en base a condición: resultado = true

Evaluación Medidas Métrica NoNulo(NroSerie)		
Regla	IdMedida	Resultado
RegNroSerie	1	true
RegNroSerie	2	false
RegNroSerie	3	false
RegNroSerie	4	true
RegNroSerie	5	true

Evaluación Reglas	
Regla	Resultado
RegFecha	1
RegNroSerie	0,6
RegNombres	0,4
RegApellidos	0,8
RegCI	0,6





Principales Etapas del Proceso

5. Medir y Evaluar Calidad de Datos

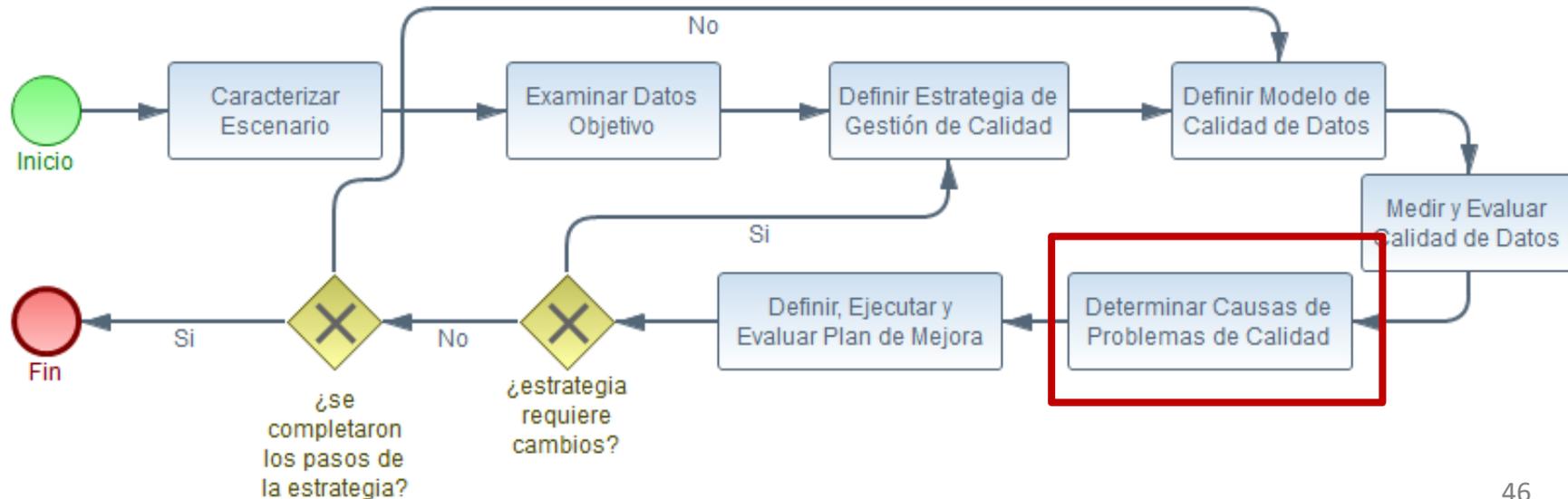
- ❑ Los principales roles del CCD involucrados en esta etapa de Medir y Evaluar Calidad de Datos son:
 - Responsable de Calidad de Datos
 - Técnico de Calidad de Datos
 - Analista de Calidad de Datos



Principales Etapas del Proceso

6. Determinar Causas de Problemas de Calidad

- El proceso consiste de siete etapas a ser llevadas a cabo por los distintos roles





Principales Etapas del Proceso

6. Determinar Causas de Problemas de Calidad

- ❑ El objetivo es determinar las causas de los problemas de calidad de datos que se confirmen o detecten en base a la evaluación de la calidad de datos

- ❑ Los resultados esperados de esta etapa son:
 - confirmación de problemas de calidad identificados
 - detección de nuevos problemas de calidad
 - determinación de causas de los problemas de calidad de datos



Principales Etapas del Proceso

6. Determinar Causas de Problemas de Calidad

- ❑ Las causas se determinan apoyándose en insumos generados en etapas anteriores:
 - resultados de las mediciones y evaluaciones
 - resultados obtenidos de la caracterización del escenario

- ❑ Ejemplo:
 - **Evaluación de Calidad:** Se comprobó que los datos de una entidad tienen problemas de completitud en un repositorio
 - **Caracterización Escenario:** Los datos de la entidad se ingresan de forma manual en una aplicación
 - **Posible Causa:** Faltan realizar controles en la aplicación de ingreso de datos



Principales Etapas del Proceso

6. Determinar Causas de Problemas de Calidad

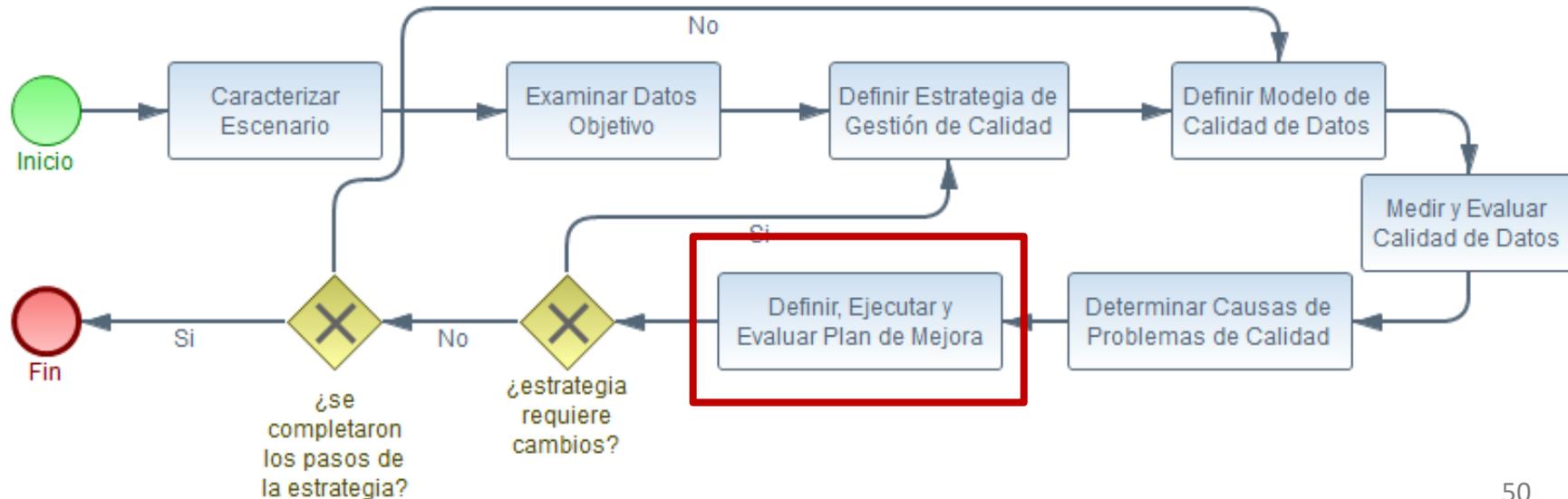
- ❑ Los principales roles del CCD involucrados en esta etapa de Determinar Causas de Problemas de Calidad son:
 - Responsable de Calidad de Datos
 - Técnico de Calidad de Datos
 - Analista de Calidad de Datos
 - Experto de Negocio
 - Experto Técnico



Principales Etapas del Proceso

7. Definir, Ejecutar y Evaluar Plan de Mejora

- El proceso consiste de siete etapas a ser llevadas a cabo por los distintos roles





Principales Etapas del Proceso

7. Definir, Ejecutar y Evaluar Plan de Mejora

- ❑ El objetivo es definir, ejecutar y evaluar un plan de mejora para la calidad de datos en el escenario de trabajo
- ❑ Los resultados esperados de esta etapa son:
 - definición de un plan de mejora de la calidad de datos
 - resultados de la ejecución y evaluación del plan de mejora



Principales Etapas del Proceso

7. Definir, Ejecutar y Evaluar Plan de Mejora

EJEMPLO

Plan de Mejora		
Acción	Descripción	Estrategia
A1	Actualizar la aplicación para que verifique que los datos de la madre nos son vacíos al momento de ingresar un certificado.	Corrección defecto
A2	Actualizar la aplicación para que verifique el formato del documento de identidad.	Corrección defecto
A3	Establecer un procedimiento (manual o automático) que permita mantener actualizadas las direcciones de correo electrónico	Re-estructura procedimiento



Principales Etapas del Proceso

7. Definir, Ejecutar y Evaluar Plan de Mejora

- ❑ Los principales roles del CCD involucrados en esta etapa de Definir, Ejecutar y Evaluar Plan de Mejora son:
 - Responsable de Calidad de Datos
 - Técnico de Calidad de Datos
 - Analista de Calidad de Datos

Temas

- ❑ Introducción
- ❑ Equipos y Roles
- ❑ Principales Etapas del Proceso
- ❑ Resumen Roles y Etapas



Resumen Roles y Etapas

Roles y Etapas

	Responsable de Calidad de Datos	Analista de Calidad de Datos	Técnico de Datos	Experto de Negocio	Experto Técnico
1a. Caracterización Técnica y de Negocio				✓	✓
1b. Caracterización de Calidad		✓		✓	
2. Examinar Datos Objetivo			✓	✓	✓
3. Definir Estrategia de Gestión de Calidad	✓	✓		✓	
4. Definir Modelo de Calidad	✓	✓		✓	
5. Medir y Evaluar Calidad	✓	✓	✓		
6. Determinar Causas Problemas	✓	✓	✓	✓	✓
7. Definir, Ejecutar y Evaluar Plan de Acción	✓	✓	✓		

Referencias

[AGE2020] AGESIC. Marco de referencia para la gestión de calidad de datos.
<https://www.gub.uy/agencia-gobierno-electronico-sociedad-informacion-conocimiento/comunicacion/publicaciones/marco-referencia-para-gestion-calidad-datos>

[Bus16] IBM Institute for Business Value. The Chief Data Officer playbook. Creating a game plan to sharpen your digital Edge. Inf. téc. IBM Institute for Business Value, 2016. URL:
<https://www.ibm.com/downloads/cas/0OGWOGAW>

[Int09] DAMA International. The DAMA Guide to the Data Management Body of Knowledge (DAMA-DMBOK). Technics Publications, 2009.

[Lee14] Yang Lee, Stuart E Madnick, Richard Y Wang, Forea Wang y Hongyun Zhang. «A cubic framework for the chief data officer: Succeeding in a world of big data».

[Wen07] Kristin Wende. «A model for data governance-Organising accountabilities for data quality management». ACIS 2007 Proceedings (2007).