

En breve iniciamos...



#PorUnPerúDigital

MESA 3: CALIDAD DE DATOS PARA LA GENERACIÓN DE VALOR PÚBLICO

Estándares, consistencia y confiabilidad de los datos



Jueves 12 de febrero



10:00 a. m.



zoom

www.gob.pe/mesasdatos



#PorUnPerúDigital

De la priorización al valor público

**MESA 1:
IDEACIÓN Y
PRIORIZACIÓN
DE DATOS DE
ALTO VALOR**

Identificando los datos que generan
mayor impacto para la ciudadanía
y el Estado.

12 de marzo 10:00 a. m.

Asiste en:

www.gob.pe/encuentros

ES POR A TU MAQUINA!

**MESA 2:
PRIORIZACIÓN
SECTORIAL
DE DATOS DE
ALTO VALOR**

Profundizando diálogos sectoriales
según las funciones y competencias
de cada sector.

13 de marzo 10:00 a. m.

Asiste en:

www.gob.pe/encuentros

ES POR A TU MAQUINA!

**MESA 3:
CALIDAD DE
DATOS PARA LA
GENERACIÓN DE
VALOR PÚBLICO**

Revisando, correlacionando y
contribuyendo a la calidad
de los datos.

12 de febrero 10:00 a. m.

Asiste en:

www.gob.pe/encuentros

ES POR A TU MAQUINA!

**MESA 4:
USO ESTRATÉGICO
DE LOS DATOS
PARA DECISIONES,
INNOVACIÓN E IA**

Del dato abierto a la evidencia,
la innovación y la inteligencia
artificial.

13 de febrero 10:00 a. m.

Asiste en:

www.gob.pe/encuentros

ES POR A TU MAQUINA!

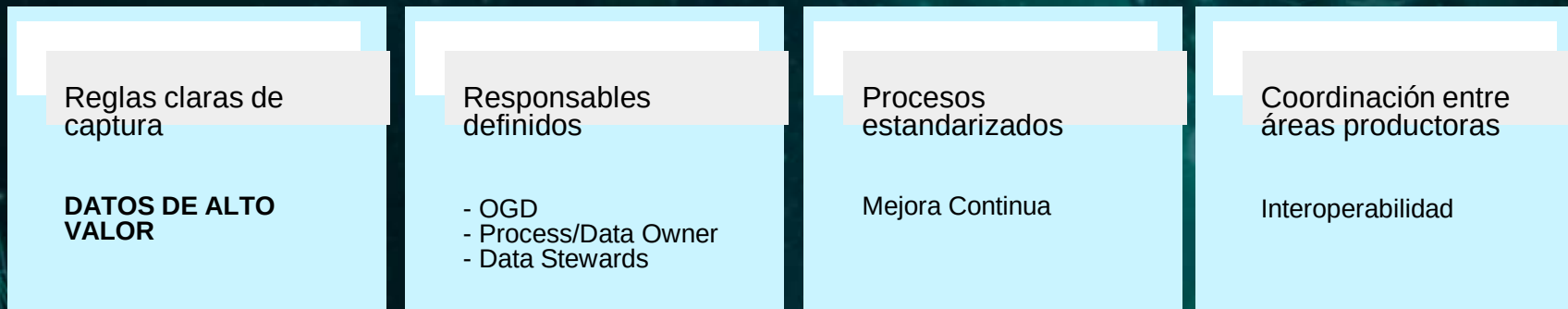


Hoy entramos al eslabón crítico que define si los datos sirven o no.

Análisis de Calidad de Datos: Mejora Continua

La calidad **no** es algo de una sola vez.

Pasar del Control Puntual a la **Mejora Continua**:



La calidad se gestiona, no se corrige al final.

Análisis de Calidad de datos: Estándares

Para que los datos GENEREN VALOR deben registrarse bajo ESTÁNDARES

(*)Estándares internacionales:

- Para DATOS : Principios FAIR



<https://www.go-fair.org/fair-principles/>

ISO/IEC 25012:2008
Modelo general de calidad de datos
→ Define 15 características clave



ISO 19157
Calidad de datos Georeferenciados

- Para Metadatos: Estándar DCAT-AP



<https://interoperable-europe.ec.europa.eu/collection/semic-support-centre/dcat-ap>

Análisis de Calidad de datos: Principios FAIR*

1. Localizable (Findable)

Para ser útiles deben ser fáciles de encontrar tanto por personas como por algoritmos.

2. Accesible

Una vez encontrados, se necesita saber cómo acceder a ellos.

3. Interoperable

Deben integrarse con otros datos y trabajar con diferentes aplicaciones.

4. Reutilizable

los datos puedan ser usados en futuras



Análisis de Calidad de datos: Metodología





Dimensiones de Calidad de los Datos



- Exactitud
- Compleitud
- Consistencia
- Integridad
- Oportunidad
- Validez
- Trazabilidad



Exactitud (Accuracy*)

Que los datos representen la realidad sin errores

Ejemplo de error:

Los beneficiarios de un bono de asistencia social al acercarse a cobrarlo presencialmente se dan cuenta que los datos de sus “nombres y apellidos” de la Lista de Beneficiarios publicada, no son los que figuran en su DNI.

Impacto:

Registrar a un ciudadano con un nombre mal escrito o que no coincida con su DNI, no permite que cobre su bono de ayuda.



Completitud (Completeness)

Que no falten datos esenciales o registros

Ejemplo de error:

El sistema de denuncias de una comisaría solicita registrar varios datos: fecha y hora, descripción del incidente, DNI del denunciante) pero el dato de “Número Telefónico” del denunciante no siempre lo solicitan en algunas comisarías

Impacto:

Expedientes de denuncias que no tienen el número de contacto del ciudadano denunciante, detiene el proceso administrativo por falta de información para poder notificarlo



Consistencia (Consistency*)

Que los datos no se contradigan cuando se comparan

Ejemplo de error:

Ante una emergencia de huaico el INDECI comparte varios datos a la ciudadanía, entre ellos por ejemplo el # de heridos en un huaico.

En paralelo el MINSA también reporta un # distinto de heridos en la misma zona por el mismo incidente.

Impacto:

Desinformación en los ciudadanos genera pánico y dificulta la asignación correcta de ayuda del Estado.





Integridad (Integrity)

Asegura que todos los datos estén debidamente vinculados

Ejemplo de error:

La base de datos de los expedientes judiciales con sentencias, se notifica a las partes involucradas sin el nombre del juez vinculado pues se borró su registro en el sistema.

Impacto:

Las sentencias quedan "huérfanas" y se puede declarar la nulidad del proceso judicial, permitiendo que criminales salgan libres.





Oportunidad (Timeliness*)

Grado en que los datos están actualizados y disponibles cuando se necesitan

Ejemplo de error:

Ocurre la desactualización de los datos de los precios de la canasta básica en el Perú, la cual se necesita para el cálculo de la inflación.

Impacto:

El BCRP toma decisiones sobre la tasa de interés con datos desactualizados, lo que afectará el bolsillo de todos los peruanos.





Validez (Validity)

Grado en que los datos cumplen con los formatos predefinidos

Ejemplo de error:

El dato del número de RUC en el portal de proveedores del Estado debe tener registrado 11 dígitos, pero una municipalidad registran los números de RUC de algunos de sus proveedores con 9 dígitos.

Impacto:

El Sistema de Pagos a Proveedores del Estado rechaza la transferencia del pago al proveedor, retrasando obras públicas o el pago a trabajadores.





Trazabilidad (Traceability)

Rastrear el origen del dato y los cambios que ha sufrido en el tiempo

Ejemplo de error:

Se modifica el presupuesto de una carretera de S/10m a S/20m, pero en los datos asociados a ese cambio, no queda registro de quién fue el funcionario que aprobó el cambio ni bajo qué sustento.

Impacto:

En las investigaciones de presuntos casos de corrupción, no se puede determinar quién cometió el presunto acto de corrupción, facilitando la impunidad en el mal uso del dinero público.





Perfilado de Datos

Es una inspección técnica inicial.

Objetivo: Obtener estadísticas descriptivas para entender qué datos hay en las tablas



Limpieza de Datos

Luego de detectados los errores, se debe corregir.

Objetivo: Obtener una base de datos lista para el uso y análisis.



Perfilado y Limpieza de Datos



PERFILADO

Detectar “valores nulos”

Detectar “valores inconsistentes”

Detectar “valores duplicados”



LIMPIEZA

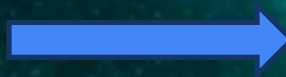
Tratamiento de nulos

Modificación de inconsistencias

Eliminación de duplicados

Perfilado y Limpieza de Datos

Detectar “valores nulos”



Tratamiento de nulos

1. Verificar Lineamientos Internos de la Entidad (p.e: “los valores vacíos son viables y se dejan en blanco, no se registran como “N/A” o “NULL”.

Peso_Neto	turno	ruta_sector	OBSERVACION
7460	T	SANTA ROSA	
6680	M	SANTA ROSA	N/A
10270	M	OLIVOS - ATALAYA	
12280	M	HUASCAR	
11800	M	HUASCAR	
10680	M	SANTA ROSA	N/A
8940	M	SANTA ROSA	
9720	M	SANTA ROSA	
16360	M	ARIAS ARAGUES-HUASCAR	falla mecanica unidad
10940	M	HUASCAR-ROSPATA	
8760	T	SAN JERONIMO - BERLIN	NULL
5900	M	SANTA ROSA	
9950	M	ARIAS ARAGUES-HUASCAR	NULL
8280	T	SANTA ROSA	

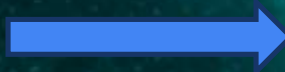
2. Los valores “N/A” y “NULL” detectados se borran y se dejan en blanco
3. Análisis de Causa Raíz

Peso_Neto	turno	ruta_sector	OBSERVACION
7460	T	SANTA ROSA	
6680	M	SANTA ROSA	
10270	M	OLIVOS - ATALAYA	
12280	M	HUASCAR	
11800	M	HUASCAR	
10680	M	SANTA ROSA	
8940	M	SANTA ROSA	
9720	M	SANTA ROSA	
16360	M	ARIAS ARAGUES-HUASCAR	falla mecanica unidad
10940	M	HUASCAR-ROSPATA	
8760	T	SAN JERONIMO - BERLIN	
5900	M	SANTA ROSA	
9950	M	ARIAS ARAGUES-HUASCAR	
8280	T	SANTA ROSA	

EJEMPLO #1

Perfilado y Limpieza de Datos

Detectar “valores nulos”



Tratamiento de nulos

1. Verificar Lineamientos Internos de la Entidad
(p.e: “los valores vacíos no deben interpretar
negación”.

2. Los valores “vacíos” deben ser
reemplazados por un “NO”
3. Análisis de Causa Raíz

FECHA_AFIILIADO	CODIGO_ANONIMIZADO	EDAD	LUBIGED	SEXO	FECHA_ULT_ATENCION	CON_DX_HIPERTENSION
20230930	8D0BAC977FE0452B850F2069284342540B3B4ED5F82D1FC30122027974F6341	64	150143	MASCULINO	20230929	
20230930	F1A577C1283FD9C96DD796F67DD3793FDDF8020C8505B1587CC4E32576825E11	51	110106	FEMENINO	20230920	
20230930	EEFEE7LB3DE1586291988F1471AF38114D2F2204E8FFEC89A663CF79E85229E	40	110207	MASCULINO	20230929	
20230930	266B3E8A0BF756FFEB6A45AC4F321F82342D641BCE3C0482E0D765CB385C4B12	60	150132	FEMENINO	20230928	SI
20230930	5044C2L2CF7EED1D8D38F13EF751A66A81DC50C6B231DC571846BFAL35B32E	48	250105	FEMENINO	20230926	
20230930	834132AF6317B674751B852F99DA7C0A2A048D6C25D6CA7E7519A369F4F0E065	67	150112	FEMENINO	20230925	
20230930	C358D28C2082F2D0FACFC669FDC540CC88809142F971B202449511FE4289D5D	32	150142	FEMENINO	20230925	
20230930	3290D077D595018C182248DF385A0522CAAFD03AE6B7CB7570520E7A0BADD09	62	100317	FEMENINO	20230925	
20230930	07B1BA0BC7969CE9A0A44BC5AE7573C9F22CEC4722B440F19FE3D05AF2C115	71	150201	MASCULINO	20230924	
20230930	86270E1B5F1958ECA98202ED29FC0AB5971D22627E8709C56055221FE78A7607	35	070101	MASCULINO	20230924	
20230930	515BF3672B9F7A5B5875BD546C2136478F51C6925C7CF15CD204C781A94DA0D	73	150101	MASCULINO	20230923	SI
20230930	C72BF7AACD09CBA642FE91B095209EB1F4548A12F967C30FA4B93E9EA9C95	56	150101	MASCULINO	20230923	
20230930	13986A413A615B18803D2465E3D30BF4B32505B2B495A17E00DFC07B5C6082D9	51	150102	MASCULINO	20230921	SI
20230930	80E57D08C5557709F878FA2A69520A5DC72672631552D6869F2E6F62F91E8A08E	37	020105	MASCULINO	20230911	

Perfilado de Datos

Detectar “valores inconsistentes”

1. Verificar Lineamientos Internos de la Entidad (p.e: cantidad máxima de alumnos será 900)
2. Verificar la consistencia de los datos (p.e: “La cantidad de alumnos nuevos debe ser un numero ENTERO entre 0 y 800”.

Modificación de inconsistencias

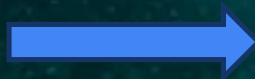
3. Los valores que no sean enteros son datos “inconsistentes” y deben modificarse.
4. Análisis de Causa Raíz.

ID	UBIGEO	COOMOD	NOMBRE	NIVEL	TIPO GESTION	DIREC	NRO. ALUMNOS
1	150119	1115286	DE TECNICAS AGROPECUARIAS-INTAP	Superior Tecnológica	Particular	CALLE LOS EUCALIPTOS MZ L LOTE 13	104
3	150119	1468255	SAN JOSE DEL SUR	Superior Tecnológica	Particular	JIRON BOLIVAR S/N MZ G LOTE 38	95
4	150119	868992	JHON F.KENNEDY	Inicial - Cuna Jardin	Particular	JIRON UNION 615	28.1
5	150119	910448	JOSE CARLOS MARIATEGUI	Inicial - Cuna Jardin	Particular	LAS VIRREYNAS MZ E LOTE 7	99
7	150119	777169	660 NIÑOS DE BELEN	Inicial - Jardin	Sector Educacion	AVENIDA 24 DE JUNIO	97.2
8	150119	777466	7104 RAMIRO PRIALE PRIALE	Inicial - Jardin	Sector Educacion	JIRON TRUJILLO S/N	30
9	150119	1071885	JUAN ENRIQUE PESTALOZZI	Inicial - Jardin	Particular	AVENIDA LAS PALMAS MZ K LOTE 19 SECTOR 8	25
10	150119	868752	SANTA ROSA	Inicial - Jardin	Particular	PASAJE LOS OLIVOS MZ D LOTE 8 SECTOR 10	10
12	150119	1055755	REYNA DE LOS ANGELES	Inicial - Jardin	Particular	MZ D LOTE 1	63
14	150119	869123	LALITO	Inicial - Jardin	Particular	CALLE LOS OLIVOS 110 FUNDO QUINTILIANA	83
16	150119	762088	657 EMILIA BARCIA BONIFATTI	Inicial - Jardin	Sector Educacion	JIRON BOLIVAR S/N MZ H2 LOTE 5A	182
18	150119	689968	7085 SANTA ROSA	Primaria	Sector Educacion	SANTA ROSA	288
19	150119	328054	6031 SANTA MARIA DE LURIN	Primaria	Sector Educacion	JIRON CASTILLA S/N	860

EJEMPLO #1

Perfilado de Datos

Detectar “valores inconsistentes”



Modificación de inconsistencias

1. Verificar Lineamientos Internos de la Entidad (p.e: las fechas deben seguir el formato aaaa.mm.dd)
2. Verificar la consistencia de los datos de datos que impliquen uso de fechas como la “fecha de nacimiento”

3. Los valores detectados debe ser modificados al formato establecido por la entidad aaaa.mm.dd.
4. Análisis de Causa Raíz.

ESCUELA_ACADEMICA	FECHA_NACIMIENTO	SEXO	ANIO EGRESO	PROMEDIO_FINAL	ANIO MATRICULA
ARQUITECTURA	1998-06-25	FEMENINO	2023	11.96	2018
ARQUITECTURA	1996-01-14	MASCULINO	2023	12.39	2018
ARQUITECTURA	1992-10-24	FEMENINO	2023	11.23	2014
ARQUITECTURA	06-05-1998	MASCULINO	2023	14.21	2017
ARQUITECTURA	1998-04-01	MASCULINO	2023	12.05	2018
DERECHO	1998-10-12	MASCULINO	2023	11.64	2017
DERECHO	1996-01-17	MASCULINO	2023	12.99	2015
ARQUITECTURA	1993-12-14	MASCULINO	2023	13.60	2013
ECONOMIA	1998-01-08	MASCULINO	2023	12.16	2018
ECONOMIA	1994-04-11	MASCULINO	2023	11.13	2012
ECONOMIA	12-12-1996	MASCULINO	2023	12.02	2015
ECONOMIA	1996-02-05	MASCULINO	2023	13.56	2016
PSICOLOGIA	1999-04-26	MASCULINO	2023	12.13	2018
PSICOLOGIA	1999-05-22	MASCULINO	2023	12.02	2018
PSICOLOGIA	1999-11-17	FEMENINO	2023	12.41	2018
PSICOLOGIA	12-08-1999	MASCULINO	2023	14.46	2019
PSICOLOGIA	1999-05-19	MASCULINO	2023	15.27	2013
PSICOLOGIA	1996-02-01	MASCULINO	2023	11.09	2015

EJEMPLO #2



Perfilado de Datos

Detectar “valores duplicados”

1. Verificar Lineamientos Internos de la Entidad (p.e: debe haber solo 1 registro por ítem de una Orden de Compra)
2. Verificar los valores duplicados en la BBDD

Eliminación de duplicados

3. Los valores detectados debe ser eliminados.
4. Análisis de Causa Raíz.

FECHA ORDEN	NRO. ORDEN	DES. GEN.	IMPORTE ÍTEM	FUENTE FINAN.	TIPO BEN.	FUNC. PROG.
20240213	000012	MUEBLES Y ENSERES	200.00	18 CANON Y SOBRECANON	B	03 PLANEAMIENTO, GESTION Y RESERVA DE CONTINGENCIA
20240223	000027	MUEBLES Y ENSERES	5,040.00	18 CANON Y SOBRECANON	B	08 COMERCIO
20240226	000039	MUEBLES Y ENSERES	25,000.00	18 CANON Y SOBRECANON	B	23 PROTECCION SOCIAL
20240228	000031	PROCESAMIENTO AUTOMATICO DE DATOS	4,671.57	18 CANON Y SOBRECANON	B	03 PLANEAMIENTO, GESTION Y RESERVA DE CONTINGENCIA
20240305	000034	PROCESAMIENTO AUTOMATICO DE DATOS	4,500.00	18 CANON Y SOBRECANON	B	23 PROTECCION SOCIAL
20240307	000051	MUEBLES Y ENSERES	1,500.00	18 CANON Y SOBRECANON	B	23 PROTECCION SOCIAL
20240307	000051	MUEBLES Y ENSERES	1,500.00	18 CANON Y SOBRECANON	B	23 PROTECCION SOCIAL
20240312	000057	PROCESAMIENTO AUTOMATICO DE DATOS	663.69	18 CANON Y SOBRECANON	B	03 PLANEAMIENTO, GESTION Y RESERVA DE CONTINGENCIA
20240312	000057	PROCESAMIENTO AUTOMATICO DE DATOS	663.69	18 CANON Y SOBRECANON	B	03 PLANEAMIENTO, GESTION Y RESERVA DE CONTINGENCIA
20240312	000057	PROCESAMIENTO AUTOMATICO DE DATOS	2,276.88	18 CANON Y SOBRECANON	B	03 PLANEAMIENTO, GESTION Y RESERVA DE CONTINGENCIA
20240312	000058	PROCESAMIENTO AUTOMATICO DE DATOS	2,204.24	18 CANON Y SOBRECANON	B	03 PLANEAMIENTO, GESTION Y RESERVA DE CONTINGENCIA
20240312	000059	PROCESAMIENTO AUTOMATICO DE DATOS	720.00	18 CANON Y SOBRECANON	B	03 PLANEAMIENTO, GESTION Y RESERVA DE CONTINGENCIA
20240312	000061	PROCESAMIENTO AUTOMATICO DE DATOS	260.00	18 CANON Y SOBRECANON	B	22 EDUCACION
20240312	000061	PROCESAMIENTO AUTOMATICO DE DATOS	260.00	18 CANON Y SOBRECANON	B	22 EDUCACION
20240312	000061	PROCESAMIENTO AUTOMATICO DE DATOS	260.00	18 CANON Y SOBRECANON	B	22 EDUCACION

Auditoria y Monitoreo

Se deben establecer
reglas automáticas que
alerten cuando los datos
"sucios" intenten entrar
al sistema.



El OGD como garante de calidad



Articular áreas
productoras

01

Definir criterios
mínimos

02

04

Asegurar
sostenibilidad

03

Priorizar datasets
críticos

Herramientas recomendadas

Nivel de Análisis de Calidad	Herramientas
Básico	Excel (Filtros, validación de datos, Power Querys).
Intermedio	Python o "R"
Profesional	Talend, SAP Data Quality, AWS Glue Databrew, etc.

<< El objetivo no es tener datos perfectos, sino datos "buenos" para generar VALOR AL CIUDADANO >>

Dinámica



REGLAS:

1. Todo el grupo participa en la dinámica.
2. Cada grupo tendrá un facilitador
3. Trabajo grupal: 45min
4. El grupo escoge 1 o 2 miembros para sustentar su trabajo.
5. Sustentación: 10min.

¿QUÉ ESPERAMOS QUE SE LLEVEN HOY?

- Importancia e impacto de un Análisis de Calidad de Datos
- Criterios mínimos para un Análisis de Calidad de Datos
- Fortalecer el Plan de Apertura 2026

CALIDAD como base para INNOVACIÓN e IA

PRIORIZACIÓN DE
DATOS

+

ANÁLISIS DE
CALIDAD

Datos de Alto
Valor
DE CALIDAD

Aplicar
Innovación e IA

Tomar
Decisiones para
agregar VALOR
PÚBLICO

¡Muchas gracias por su participación!

contacto@datosabiertos.gob.pe



PERÚ

Presidencia
del Consejo de Ministros

¡EL PERÚ A TODA
MÁQUINA!