



INSTITUTO NACIONAL DE ESTADISTICA

Los metadatos estadísticos

Departamento de metodología y desarrollo de la producción
estadística

Madrid, Junio de 2016

Ana Isabel Sánchez-Luengo Murcia
nomenclaturas@ine.es



Los metadatos estadísticos

- Definición de los metadatos
- Contexto internacional
- Los metadatos y sus tipos
- Estándares y metadatos
- Implementación de un sistema de metadatos

¿Qué es un metadato?



Documentar

¿Qué es un metadato?

Seguridad alimentaria: Ejemplo de metadatos orientado a lo usuarios o clientes



Campaña publicitaria: "La etiqueta lo dice todo" Gobierno de España

¿Qué es un metadato?

ODE1344461

- ➡ Primer dígito: modo de cría de las gallinas
 - Si es un 3 "criadas en jaulas":
 - Si es un 2 "criadas en suelo sólo gallinero":
 - Si es un 1 "camperas": Aire libre
 - Si es un 0 "de producción ecológica, incluida alimentación
- ➡ Dos letras siguientes: Estado Miembro de producción
- ➡ Resto de dígitos: Granja de producción
- ◆ En la imagen el código del huevo 0-DE-1344461 indica que es un Bio-huevo, de Alemania región 13 Mecklenburg-Vorpommern granja 1344461.
- ◆ El tamaño del huevo va en la caja de fuera XL, L ...



¿Qué es un metadato?



0034 ¿Qué significa?

Crecimiento del PIB

¿De dónde?

De España

¿En qué periodo?

2006

¿Cuál es el dato?

3,4

¿En qué unidad esta el dato?

En
porcentaje

Fuente **INE**



¿Qué es un metadato?

COLORES

Los nuevos métodos de trabajo y comunicaciones han hecho necesaria una profunda reforma en el edificio que, desde 1972, albergaba la sede del INE. Además de incorporar las últimas tecnologías y diseños en su concepción arquitectónica, el elemento más llamativo del nuevo edificio es la decoración de su fachada, obra de Pepe Cruz Novillo, artista plástico que lleva varios años desarrollando su concepto de "diagrama" sobre diversos formatos y soportes. La característica común de estas obras es la combinación de un número variable de elementos monocromos, sonoros, fotográficos o tridimensionales.

El sistema de transcripción de cifras a colores propuesto por Cruz Novillo es condso: cada color equivale a un dígito del 0 al 9 y se representa mediante rectángulos monocromáticos que se articulan con la modulación de los antepechos de las ventanas del edificio.

Asimismo, cada cifra —con su color asociado— se corresponde con una nota musical.

Se han seleccionado 58 Indicadores estadísticos para, a partir de ellos, desarrollar los 58 *diagramas decafonicos* de dígitos que cubren en sentido horizontal y descendente, planta por planta, las fachadas Sur, Este y, parcialmente, Norte del edificio.

Para una institución como el INE, que difunde los resultados de sus estadísticas principalmente en forma numérica, resulta especialmente adecuado el concepto propuesto por el artista.



Los 58 indicadores

001 Densidad de España en m²	004-86	029 Número de hogares (2004, Ter.)	14.405.940
002 Población real densa (1 enero 2005)	44.108.330	030 Tasa de mortalidad del hogar en número de personas (2004, Ter.)	3,91
003 PIB por cápita en precios constantes en euros (2004)	19.637	031 Hogares unipersonales de porcentaje (2004, Ter.)	14,59
004 Población real densa, mujeres (1 enero 2005)	22.073.881	032 Hogares con solo a más personas en porcentaje (2004, Ter.)	2,93
005 Población real densa, hombres (1 enero 2005)	21.709.689	033 PIB precios constantes en millones de euros (2004)	927.346
006 Población real densa, extranjeros (1 enero 2005)	2.052.010	034 Crecimiento del PIB en España en porcentaje (2005)	2,40
007 Población real densa, mayores 65 años en porcentaje (1 enero 2005)	16,82	035 Crecimiento del PIB en UE-25 en porcentaje (2005)	1,00
008 Número medio de hijos por mujer en España (2004)	1,54	036 PIB en paridad de poder adquisitivo UE-25 100. España (2004)	97,80
009 Población real densa, menores 15 años en porcentaje (1 enero 2005)	1,40	037 Contribución contributiva (2005)	137.840
010 Población real densa, menores 15 años en porcentaje (1 enero 2005)	21.159.012	038 Número de empresas (2005)	3.094.139
011 Población real densa (2005, 4º t.)	18.214.257	039 Empresas sin actividad en porcentaje (2005)	51,37
012 Población real densa (2005, 4º t.)	1.941.235	040 Suplantación de bienes en miles de euros (2005)	153.559.855
013 Tasa de paro en España (2005, 4º t.)	9,30	041 Suplantación a la UE-25 en porcentaje (2005)	72,58
014 Tasa de paro en UE-25 (2005, 4º t.)	9,30	042 Importación de bienes en miles de euros (2005)	221.071.586
015 Alumnos matriculados en enseñanzas obligatorias menores de 16 años (curso 2004/2005)	57.76.822	043 Deficit de la balanza comercial en miles de euros (2005)	77.512.317
016 Alumnos matriculados en bachillerato FP (curso 2004/2005)	1.154.424	044 Turismo y viajes, ingresos en miles de euros (2005)	37.762.738
017 Alumnos matriculados en la universidad (curso 2004/2005)	1.403.287	045 Turismo y viajes, pagos en miles de euros (2005)	12.125.215
018 Número de profesores no universitarios (curso 2004/2005)	599.270	046 Importación de la UE-25 en porcentaje (2005)	65,62
019 Número de profesores universitarios (curso 2004/2005)	101.469	047 Deficit de la balanza corriente en miles de euros (2005)	68.562.142
020 Producción editorial en número de ejemplares en miles (2005)	321.005	048 Tasa de ahorro sobre la renta disponible en porcentaje (2004)	11,30
021 Gasto público en educación en España en % del PIB (2005)	4,40	049 Múltiplos (2005)	1.763.847
022 Gasto público en educación en UE-25 en % del PIB (2005)	5,40	050 Viviendas familiares (2001)	23.945.954
023 Número de médicos (2004)	144.698	051 Viviendas colectivas (2001)	11.446
024 Número de profesores sanitarios (2004)	524.769	052 Formas de sujeción estadísticas en miles (2004)	24.425.04
025 Número de hospitales (2005)	779	053 Viviendas en alquiler en España en porcentaje (2001)	15,10
026 Número de camas por cada 1.000 habitantes (2005)	2,81	054 Viviendas en alquiler en UE-25 en porcentaje (2001)	35,10
027 Índice de Precios de Consumo en España. % crecimiento (variación 2005/2004)	2,40	055 Viviendas con ordenador (2001)	7.709.342
028 Índice de Precios de Consumo Aumentado en UE-25. % crecimiento (variación 2005/2004)	2,30	056 Empresas con ordenador en porcentaje (2004/2005)	97,40
029 Índice de Precios de Consumo Aumentado en UE-25. % crecimiento (variación 2005/2004)	2,30	057 Empresas con internet en porcentaje (2004/2005)	98,55
030 Índice de Precios de Consumo Aumentado en UE-25. % crecimiento (variación 2005/2004)	2,30	058 Viviendas con teléfono móvil en miles (2005)	12.256

Base: recopilada en marzo de 2005



Diagrama decafonico de dígitos

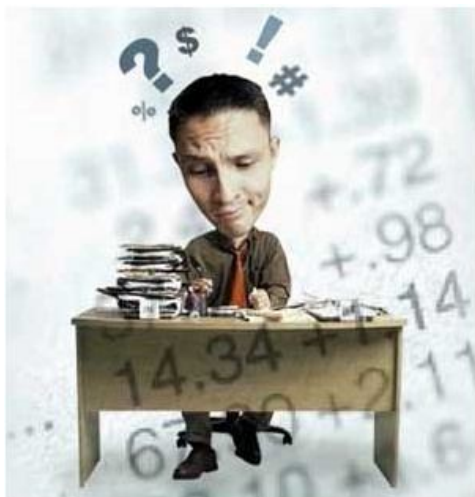
Correspondencia entre colores, dígitos y notas musicales

1. Do	2. Re bemol Do sostenido	3. Re	4. Mi bemol Re sostenido	5. Mi
6. Fa	7. Sol bemol Fa sostenido	8. Sol	9. La bemol Sol sostenido	0. Si
Indicador estadístico	Número entero 2 X 1	Número decimal 1 X 1		

Ejemplo:
052 Parque de automóviles matriculados en miles (2004): 26.432.64



El proceso estadístico y los metadatos



NUESTRO NEGOCIO

**LA PRODUCCION
ESTADISTICA**

NUESTRO PRODUCTO

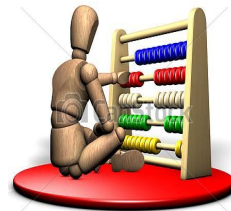
LOS DATOS



Reconocimiento de la industria de la “estadística oficial”

[illegible]

Un nuevo modelo de producción



Folio 8 verso

Insigne della *Repubblica* *di* *San Marino*
 Confessione di *S. Maria* *di* *San Marino*
 Impero di *San Marino*

	1870	1880	1890
Popolazione	1.257	1.298	1.851
Terreno	100	100	100
Grano	7	7	7
Industria	7	7	7
Artigianato	7	7	7
Commercio	7	7	7
Finanze	7	7	7
Religione	7	7	7
Scienze	7	7	7
Lettere	7	7	7
Giurisprudenza	7	7	7
Medicina	7	7	7
Letteratura	7	7	7
Arte	7	7	7
Scienze	7	7	7
Lettere	7	7	7
Giurisprudenza	7	7	7
Medicina	7	7	7
Letteratura	7	7	7
Arte	7	7	7
Scienze	7	7	7
Lettere	7	7	7
Giurisprudenza	7	7	7
Medicina	7	7	7
Letteratura	7	7	7
Arte	7	7	7
Scienze	7	7	7
Lettere	7	7	7
Giurisprudenza	7	7	7
Medicina	7	7	7
Letteratura	7	7	7
Arte	7	7	7
Scienze	7	7	7
Lettere	7	7	7
Giurisprudenza	7	7	7
Medicina	7	7	7
Letteratura	7	7	7
Arte	7	7	7
Scienze	7	7	7
Lettere	7	7	7
Giurisprudenza	7	7	7
Medicina	7	7	7
Letteratura	7	7	7
Arte	7	7	7
Scienze	7	7	7
Lettere	7	7	7
Giurisprudenza	7	7	7
Medicina	7	7	7
Letteratura	7	7	7
Arte	7	7	7
Scienze	7	7	7
Lettere	7	7	7
Giurisprudenza	7	7	7
Medicina	7	7	7
Letteratura	7	7	7
Arte	7	7	7
Scienze	7	7	7
Lettere	7	7	7
Giurisprudenza	7	7	7
Medicina	7	7	7
Letteratura	7	7	7
Arte	7	7	7
Scienze	7	7	7
Lettere	7	7	7
Giurisprudenza	7	7	7
Medicina	7	7	7
Letteratura	7	7	7
Arte	7	7	7
Scienze	7	7	7
Lettere	7	7	7
Giurisprudenza	7	7	7
Medicina	7	7	7
Letteratura	7	7	7
Arte	7	7	7
Scienze	7	7	7
Lettere	7	7	7
Giurisprudenza	7	7	7
Medicina	7	7	7
Letteratura	7	7	7
Arte	7	7	7
Scienze	7	7	7
Lettere	7	7	7
Giurisprudenza	7	7	7
Medicina	7	7	7
Letteratura	7	7	7
Arte	7	7	7
Scienze	7	7	7
Lettere	7	7	7
Giurisprudenza	7	7	7
Medicina	7	7	7
Letteratura	7	7	7
Arte	7	7	7
Scienze	7	7	7
Lettere	7	7	7
Giurisprudenza	7	7	7
Medicina	7	7	7
Letteratura	7	7	7
Arte	7	7	7
Scienze	7	7	7
Lettere	7	7	7
Giurisprudenza	7	7	7
Medicina	7	7	7
Letteratura	7	7	7
Arte	7	7	7
Scienze	7	7	7
Lettere	7	7	7
Giurisprudenza	7	7	7
Medicina	7	7	7
Letteratura	7	7	7
Arte	7	7	7
Scienze	7	7	7
Lettere	7	7	7
Giurisprudenza	7	7	7
Medicina	7	7	7
Letteratura	7	7	7
Arte	7	7	7
Scienze	7	7	7
Lettere	7	7	7
Giurisprudenza	7	7	7
Medicina	7	7	7
Letteratura	7	7	7
Arte	7	7	7
Scienze	7	7	7
Lettere	7	7	7
Giurisprudenza	7	7	7
Medicina	7	7	7</



Usuarios de los metadatos

USUARIOS EXTERNOS

- ✓ Prensa
- ✓ Académicos
- ✓ Políticos
- ✓ Particular

USUARIOS INTERNOS

- ✓ Órganos de decisión
- ✓ Gestión y planificación
- ✓ Elaboradores de información



Principal reto

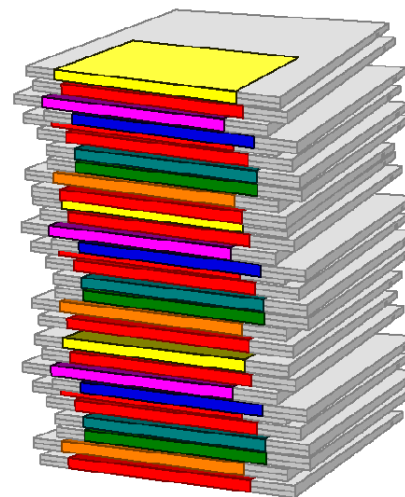
- Aumento ingente de la información disponible:
 - Fuentes administrativas
 - Otros tipo de ficheros
 - Internet
- ❖ ¿Tenemos capacidad para gestionarlo?

¿Para qué sirven los metadatos?



Documentar

Estructurar



Conseguir un proceso

EFICIENTE y de **CALIDAD**

Definición de metadatos

Los **metadatos** se definen como la información necesaria para el uso e interpretación de las estadísticas. Los metadatos describen los datos con las definiciones de las poblaciones, las variables, la metodología y la calidad.

Definición basada en : WALLGREN, Anders and Britt (Statistics Sweden), "Register-based Statistics. Administrative Data for Statistical Purposes", John WILEY & Sons, Ltd, England, 2007

Fuente: Eurostat, "Terminology relating to the Implementation of the Vision on the Production Method of EU Statistics", W. Radermacher, A. Baigorri, D. Delcambre, W. Kloek, H. Linden, European Conference on Quality in Statistics 2010 (Q 2010), Luxembourg, 2010

Los metadatos estadísticos

- Definición de los metadatos
- Contexto internacional
- Los metadatos y sus tipos
- Estándares y metadatos
- Implementación de un sistema de metadatos

Un poco de historia

1983 nacimiento de Internet

1998 creación del grupo METIS

Metadatos para la documentación

Metadatos para el proceso

2000 creación de INEbase



Grupo METIS

- Grupo de trabajo origen de los metadatos:
 - Empieza a finales de los 90
 - Poca presencia de la estadística en los buscadores de Internet
 - Interés de una mejor documentación en la difusión

Common Metadata Framework (CMF)

- Parte A: Metadatos Estadísticos en el contexto corporativo.
- Parte B: Conceptos, estándares, modelos y registros.
- Parte C: Modelo genérico de producción estadística: GSBPM
- Parte D: Implementación

<http://www1.unece.org/stat/platform/display/metis/The+Common+Metadata+Framework>

- **Reglamento** coordinar todos los metadatos
 - ✓ Tipos de metadatos
 - ✓ Intercambio de datos y metadatos

Los metadatos y la calidad

Definiciones:
A efectos del presente documento

- Por estadísticas europeas se entenderán las estadísticas comunitarias, tal como se definen en el Reglamento (CE) n° 322/97 del Consejo, de 17 de febrero de 1997, sobre la estadística comunitaria, elaboradas y definidas por las autoridades nacionales de estadística y la autoridad estadística de la Comunidad (Eurostat) de conformidad con el artículo 236, apartado 2, del Tratado.
- Por autoridad estadística se entenderá, a escala nacional, el Instituto Nacional de Estadística (INE) y los demás organismos estadísticos responsables de elaborar y difundir las estadísticas europeas y, a escala europea, Eurostat.
- Por Sistema estadístico europeo, en lo sucesivo denominado SEE, se entenderá la asociación formada por Eurostat, los institutos nacionales de estadística y los demás organismos estadísticos responsables, en cada Estado miembro, de elaborar y difundir las estadísticas europeas.




Código de buenas prácticas de las estadísticas europeas

Para las autoridades estadísticas nacionales y comunitarias

Adoptado por el Comité del Programa Estadístico del 24 de febrero de 2005 y publicado en la Recomendación de la Comisión, de 25 de mayo de 2005, sobre la independencia y responsabilidad de las autoridades estadísticas nacionales y comunitarias.



El Código de conducta se basa en quince Principios. Las autoridades encargadas de la gobernanza y las autoridades estadísticas de la Unión Europea se comprometen a respetar los principios establecidos en el presente Código y a revisar su aplicación periódicamente utilizando los Indicadores de buenas prácticas correspondientes a cada uno de los quince principios, que deberán utilizarse como referencia.

Los metadatos y la calidad

Principio 15. Accesibilidad y claridad

Las estadísticas europeas deberían presentarse de forma clara y comprensible, difundirse de forma adecuada y conveniente y estar disponibles, asimismo se debería permitir el acceso a las mismas de forma imparcial, con metadatos y orientación de apoyo.

Indicadores

- Las estadísticas se presentan de tal forma que facilitan una interpretación adecuada y comparaciones significativas.
- Los servicios de difusión utilizan una tecnología moderna de información y comunicación y, si procede, una copia impresa tradicional.
- Cuando sea posible se suministren análisis a medida y se hacen públicos.
- Se pueda permitir el acceso a los microdatos a efectos de investigación. Dicho acceso está sometido a protocolos estrictos.
- Los metadatos están documentados con arreglo a sistemas de metadatos normalizados.
- Se mantiene informado a los usuarios sobre la metodología de los procesos estadísticos y la calidad de la producción estadística respecto a los criterios de calidad del SEE.

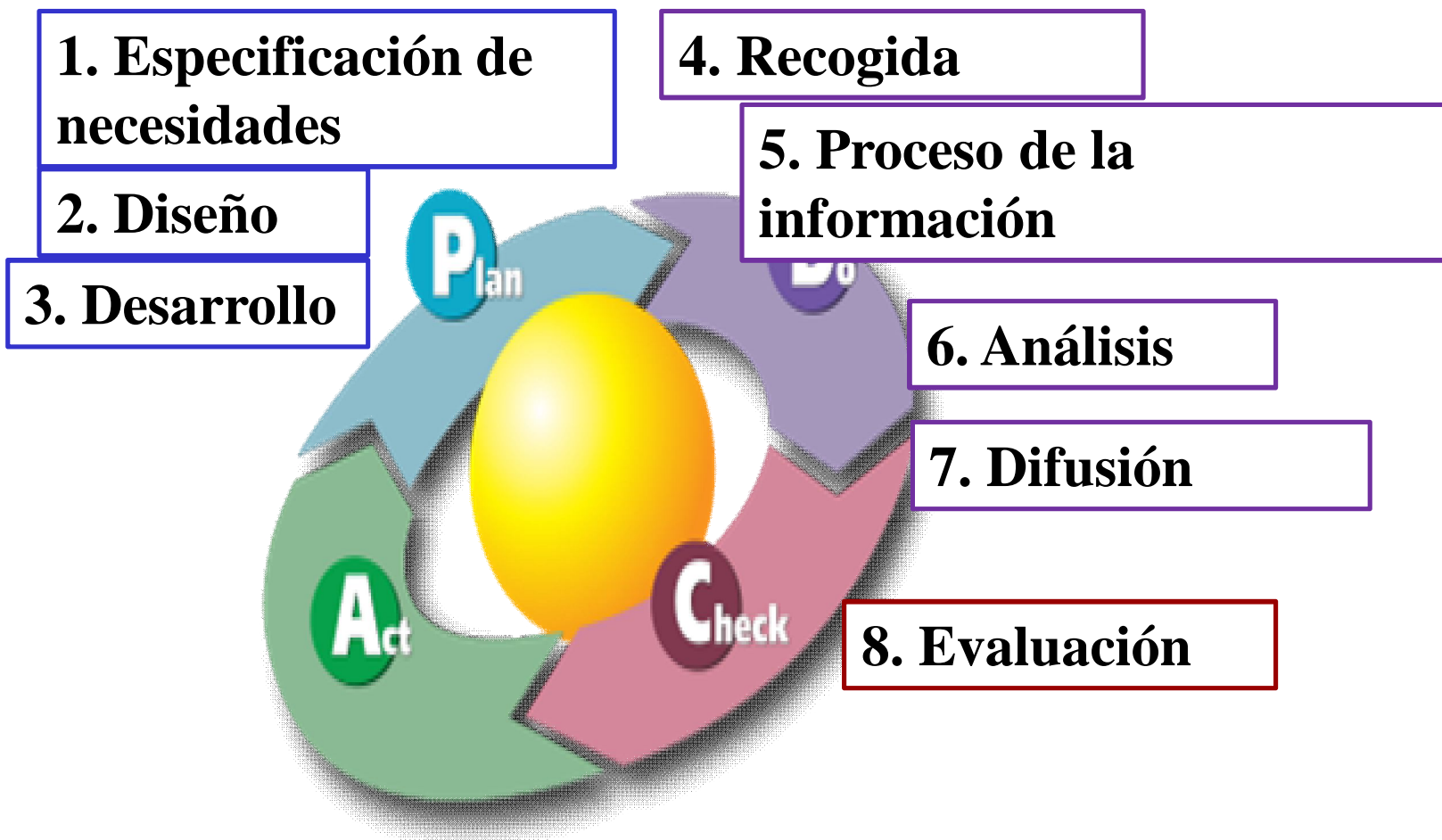
Definición de metadatos

Los **metadatos** se definen como la información necesaria para el uso e interpretación de las estadísticas. Los metadatos describen los datos con las definiciones de las poblaciones, las variables, la metodología y la calidad.

Definición basada en : WALLGREN, Anders and Britt (Statistics Sweden), "Register-based Statistics. Administrative Data for Statistical Purposes", John WILEY & Sons, Ltd, England, 2007

Fuente: Eurostat, "Terminology relating to the Implementation of the Vision on the Production Method of EU Statistics", W. Radermacher, A. Baigorri, D. Delcambre, W. Kloek, H. Linden, European Conference on Quality in Statistics 2010 (Q 2010), Luxembourg, 2010

El proceso estadístico



Los metadatos estadísticos

- Definición de los metadatos
- Contexto internacional
- **Los metadatos y sus tipos**
- Estándares y metadatos
- Implementación de un sistema de metadatos

Tipos de metadatos

CALIDAD

REFERENCIA

- IOE
- Metodologías estandarizadas

Operación Estadística

PROCESO

• Procesos Imputación, Codificación,...

Proceso

ESTRUCTURA

- Clasificaciones
- Unidades Estadísticas
- Conceptos

Bases de Datos



Los metadatos estadísticos



Proyectos



Metadatos
de referencia



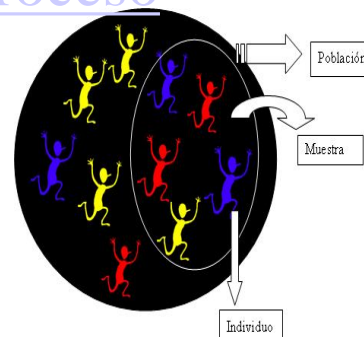
Metadatos
de proceso



**Clasificacio
nes**



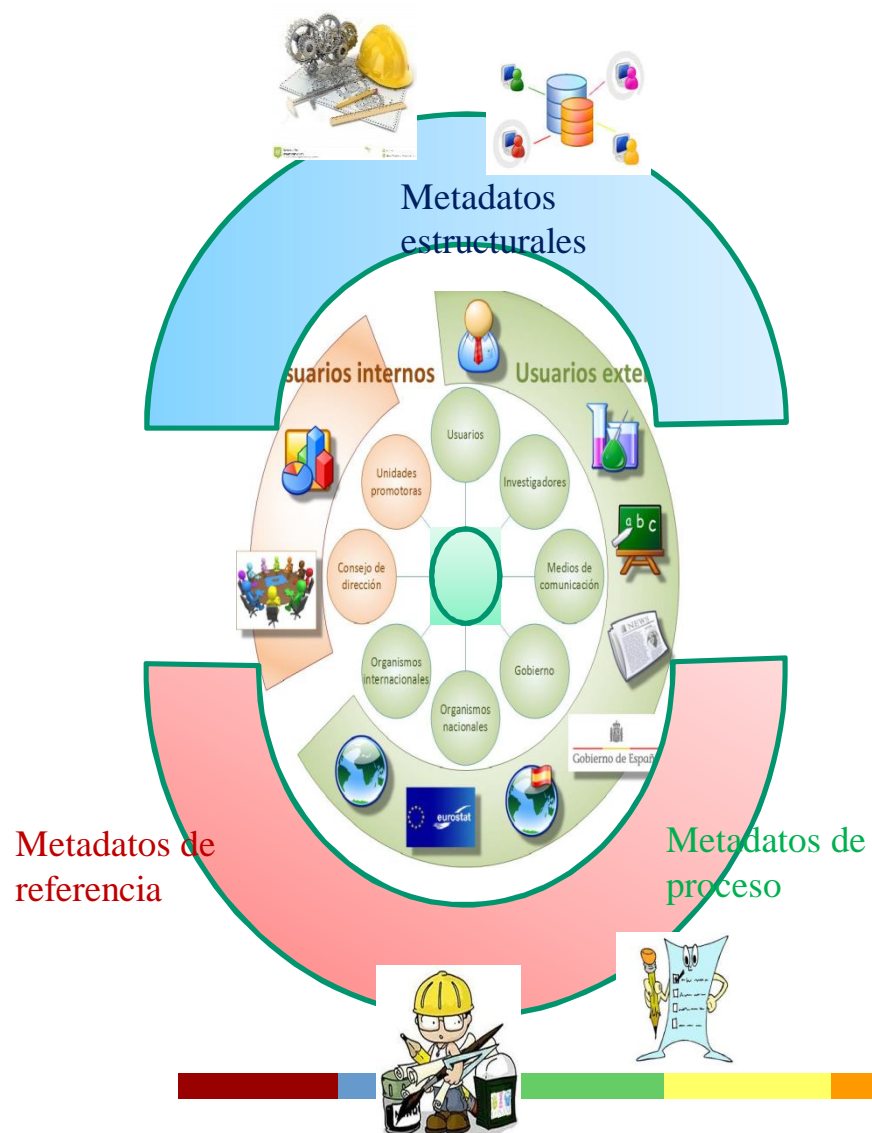
DEFine:
Conceptos



**Unidades
estadísticas**



Usuarios



Definición de metadatos

Metadatos de referencia:

Metadatos describiendo el contenido y la calidad de los datos estadísticos y el proceso de producción estadística.

Fuente: Eurostat



Los metadatos estadísticos

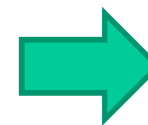


Definición de metadatos

Los **metadatos estructurales** se utiliza para identificar, describir formalmente o recuperar datos estadísticos.

Por ejemplo, los metadatos estructurales incluyen los títulos de las variables, así como las unidades empleadas, las clasificaciones utilizadas, etc.

Fuente: Eurostat



Los metadatos estadísticos



Los metadatos estadísticos

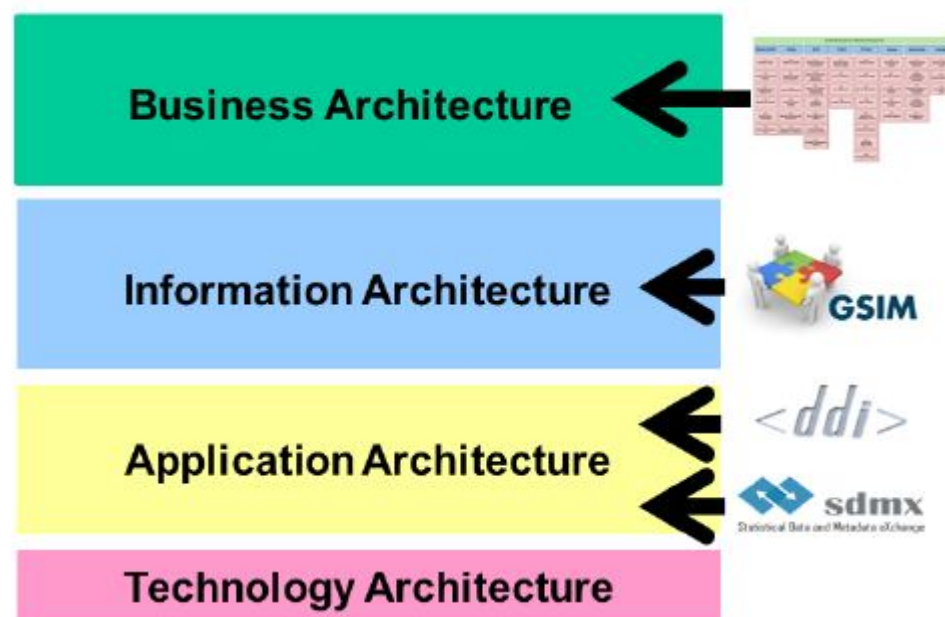
- Definición de los metadatos
- Contexto internacional
- Los metadatos y sus tipos
- **Estándares y metadatos**
- Implementación de un sistema de metadatos

Estándares internacionales para metadatos

- ◆ GSBPM (Generic Statistical Business Process Model)
- ◆ DDI (Data Documentation Initiative)
- ◆ GSIM (Generic Statistical Information Model)
- ◆ SDMX (Statistical Data and Metadata eXchange)

Estándares internacionales para metadatos

Levels of architecture



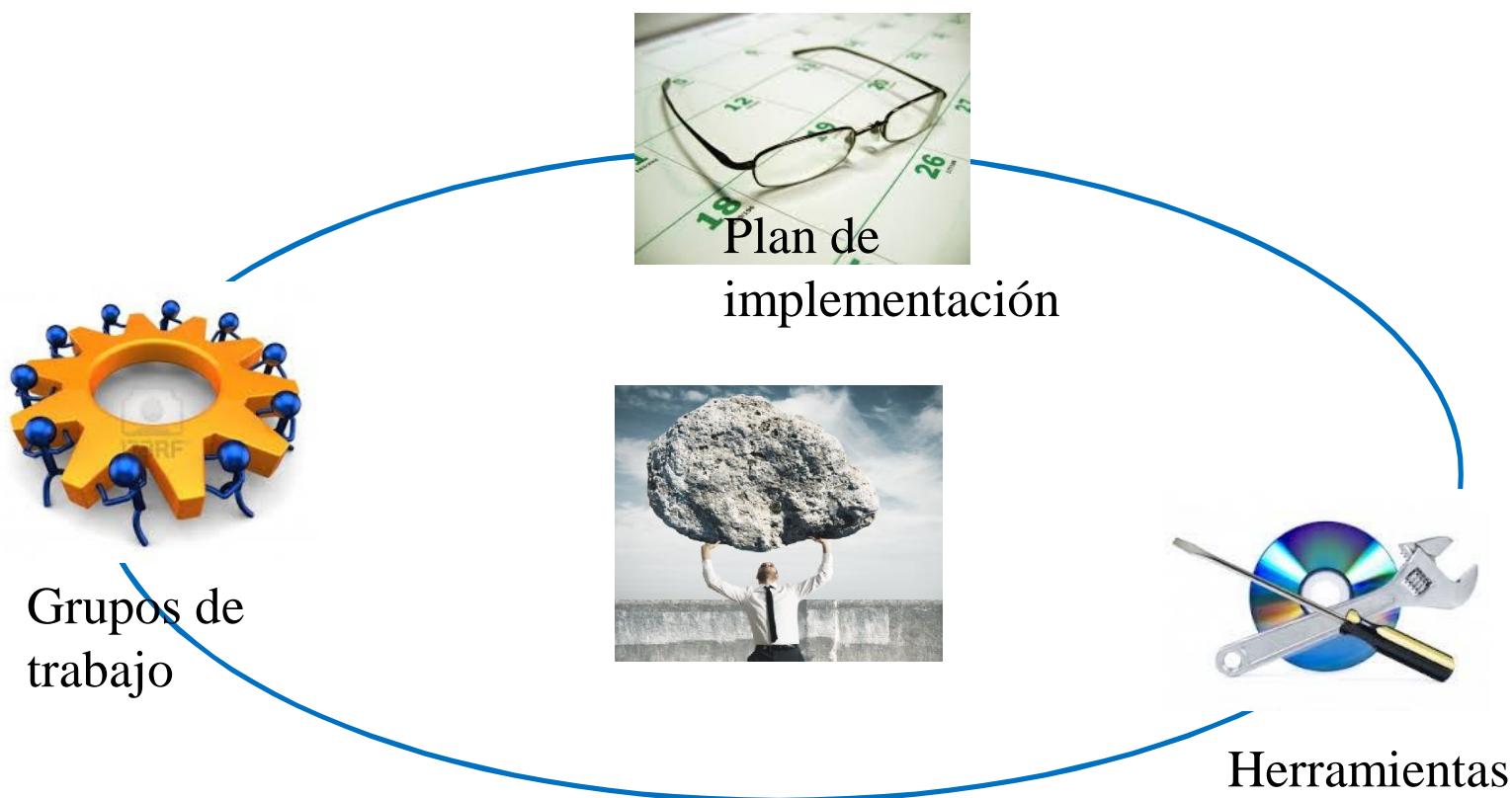
Los metadatos estadísticos

- Definición de los metadatos
- Contexto internacional
- Los metadatos y sus tipos
- Estándares y metadatos
- Implementación de un sistema de metadatos

Implementación de sistemas de metadatos



Implementación de sistemas de metadatos



Implementación de sistemas de metadatos

- ➡ Eficiencia: Reutilización (procesos, ...) e integrabilidad
- ➡ Transparencia
- ➡ Desarrollo sistemático
- ➡ Sencillez
- ➡ Amigabilidad

Implementación de sistemas de metadatos

- ➔ Debe ser un modelo que cumpla:
- ◆ Bien explicado – difusión o comunicabilidad
 - ◆ Estable en el tiempo
 - ◆ Aplicable de una forma segura
 - ◆ Completo para todo el proceso
 - ◆ Compartido: Listo para ser usado
 - ◆ Sostenible
 - ◆ Metadato unido al “dato”
 - ◆ Los metadatos han de estar vinculados a otras aplicaciones de la Institución

Lecciones aprendidas

INSTITUTO NACIONAL DE ESTADISTICA

IMPLEMENTACION

Información
necesaria para
entender los datos
y hacer eficiente el
proceso

Junto a los
datos

¿Qué son?

¿Dónde?

¿Cuándo?

En todas las fases del
proceso

¿Quién? ¿Cómo?

¿Por
qué?

Todas las
unidades la
institución

Apoyo
institucional

Documentar
Estructurar
Reutilizar
Estandarizar

Los metadatos estadísticos



Conclusiones



- ✓ Documentar
- ✓ Normalizar
- ✓ Reutilizar



- ✓ Información de mayor calidad



- ✓ Proceso más eficiente



**Servicio de postventa:
anaisabel.sanchez.luengo@ine.es**

