

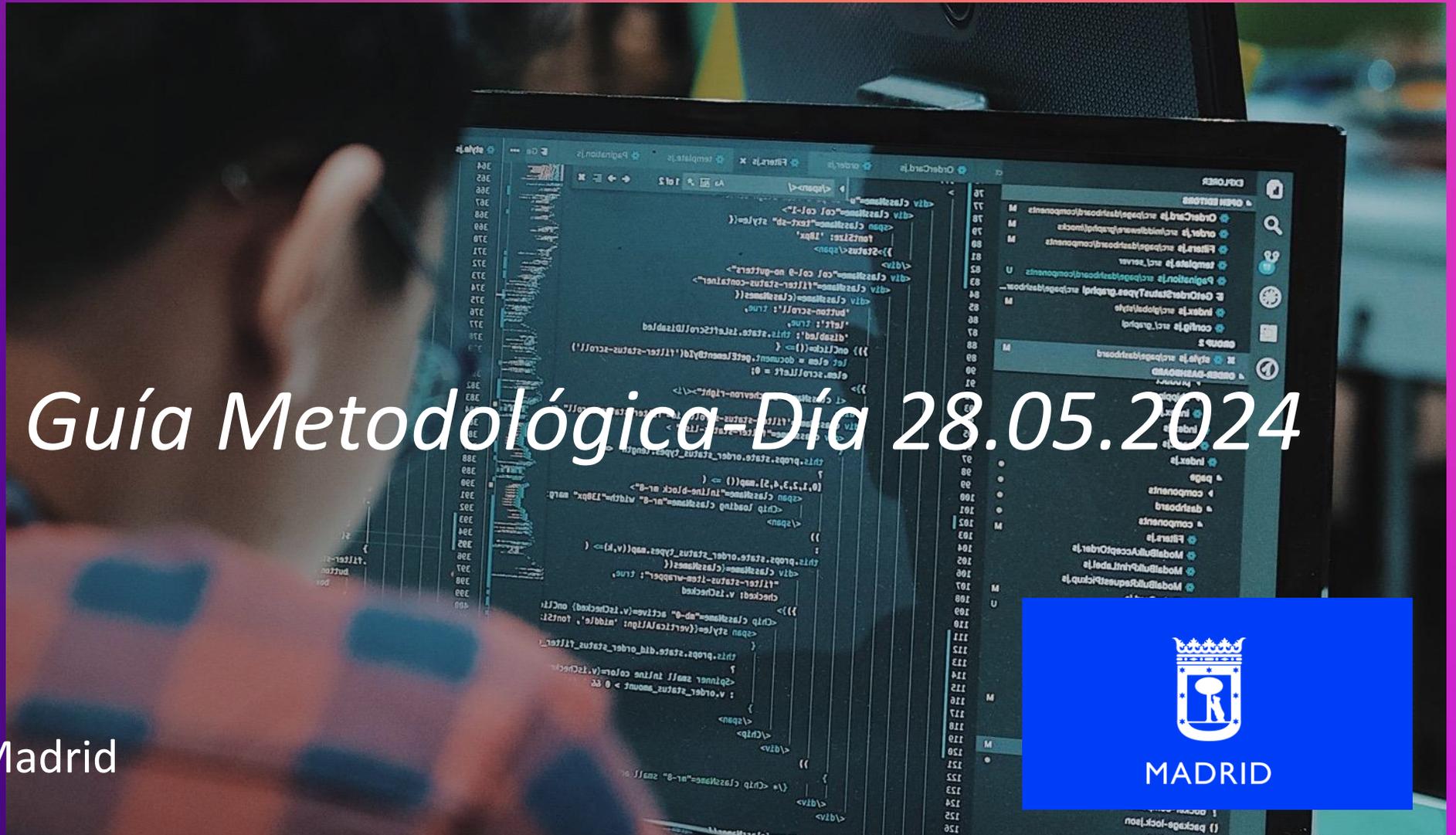
Sesión calidad de los datos abiertos

Proyecto Perfilado de Datos

Formación *Guía Metodológica* Día 28.05.2024

Ayuntamiento de Madrid

Accenture



INDICE

1. INTRODUCCIÓN Y OBJETIVOS
2. AGENDA FORMACIÓN
3. MEJORES PRÁCTICAS PARA LA PUBLICACIÓN DE DATOS ABIERTOS
 - a) METADATOS
 - b) FORMATOS REUTILIZABLES
 - c) DATOS
4. ERRORES MÁS COMUNES DETECTADOS EN DATOS.MADRID.ES
 - a) METADATOS
 - b) FORMATOS REUTILIZABLES
 - c) DATOS
5. GUÍAS EXTERNAS
6. FORMACIÓN BÁSICA PYTHON
7. DUDAS Y ENCUESTA FINAL



INTRODUCCIÓN Y OBJETIVOS



1. INTRODUCCIÓN Y OBJETIVOS

¿Qué son los datos abiertos?

Los datos abiertos deben estar **disponibles** para que cualquier persona los **use, modifique y comparta libremente, sin restricciones de copyright, patentes u otros mecanismos de control**. Están disponibles para el público en general **de forma gratuita**. Además, destacan por estar al **máximo nivel de detalle** y su capacidad para proporcionar información precisa y completa que pueda ser utilizada de manera efectiva por una amplia variedad de usuarios y aplicaciones.

Principios de los datos abiertos:

1. Completos
2. Primarios
3. Oportunos
4. Accesibles
5. Procesables por máquinas
6. No discriminatorios
7. No propietarios
8. Licencia libre

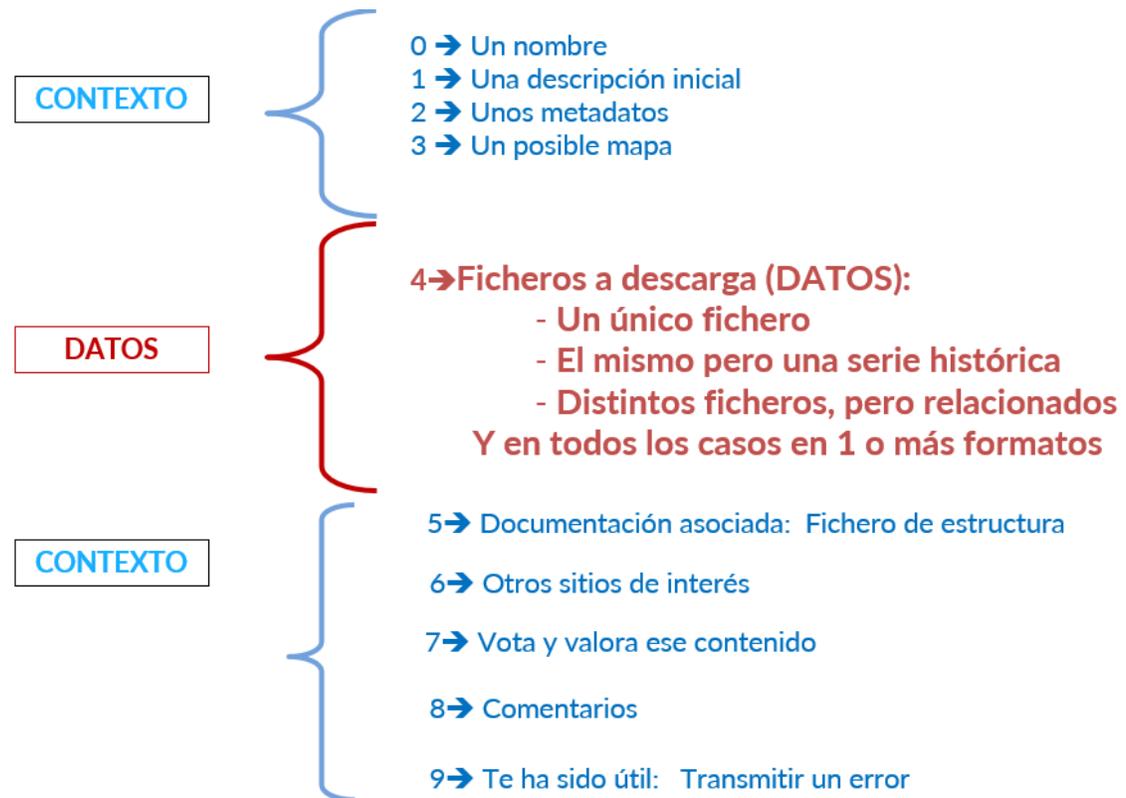
¿Por qué es necesaria la calidad de los datos?

Los datos precisos y confiables son indispensables para la toma de **decisiones informadas** y la **generación de análisis confiables**. Además, los datos de alta calidad son relevantes y útiles para una amplia gama de usuarios y aplicaciones, ya que proporcionan información precisa y detallada que puede ser utilizada de manera efectiva para **resolver problemas y generar conocimiento en diversos campos**.

Deben ser exactos, completos, consistentes y actualizados.

1. INTRODUCCIÓN Y OBJETIVOS

¿Qué estructura tiene un conjunto de datos del portal?



1. INTRODUCCIÓN Y OBJETIVOS

Para las **administraciones públicas**, **datos abiertos** son fuentes de **importante valor**, ya que fomentan la transparencia y participación ciudadana, la innovación y la mejora de servicios públicos.

A **partir de 2010**, se observó una **rápida expansión** de portales de datos abiertos alrededor del mundo, lo que ha implicado diferentes **estándares y mejores prácticas** alrededor de los mismos.

En la literatura sobre datos abiertos existen diferentes enfoques sobre cuáles son los atributos y características que los conjuntos de datos deben tener para ser considerados de calidad. Algunas de las **iniciativas** más **importantes** son:

Carta Internacional de los Datos Abiertos
(["opendatacharter.net"](http://opendatacharter.net))

Guía práctica mejora de la calidad de Datos Abiertos
(["datos.gob.es/es/documentacion"](http://datos.gob.es/es/documentacion))



El objetivo de la sesión de **formación** de la **Guía Metodológica** será ayudar **al Ayuntamiento de Madrid** y sus fuentes de datos a mejorar sus controles y validaciones de **calidad, utilidad, transparencia y reutilización de sus datos** dentro del marco del **Proyecto de Perfilado de Datos**

1. INTRODUCCIÓN Y OBJETIVOS

Beneficios internos (para el Ayuntamiento)

Portal de Datos Abiertos

Aumenta la calidad de los datos

Igualdad de oportunidades de acceso
(Neutralidad = Limpieza)

Se reducen solicitudes de información

Se evita pérdida de datos

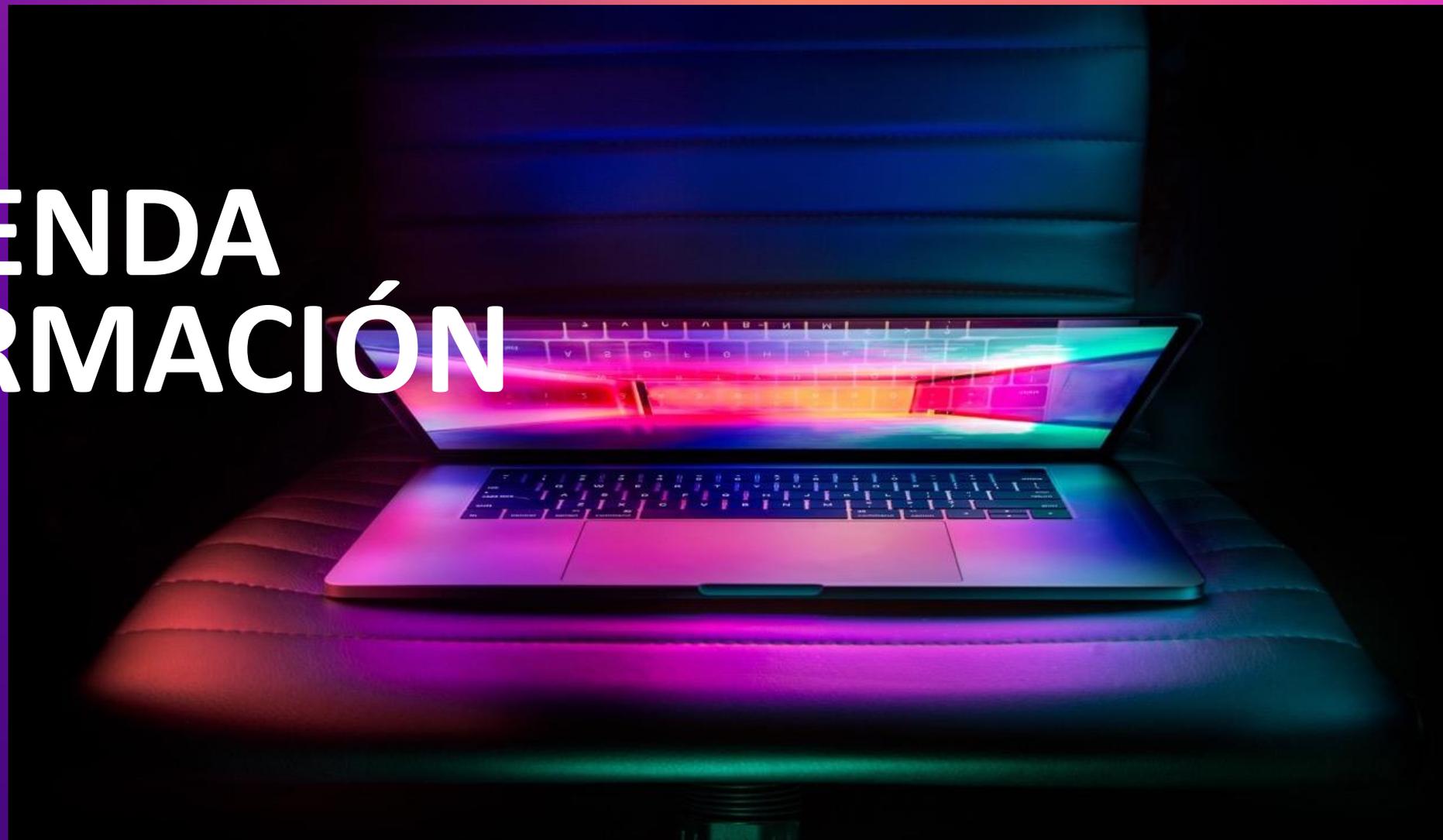
Mejora la toma de decisiones

Aumenta la coordinación, conocimiento y USO
POR OTRAS UNIDADES

Se ahorra en desarrollo de aplicaciones

Pone en valor vuestros datos

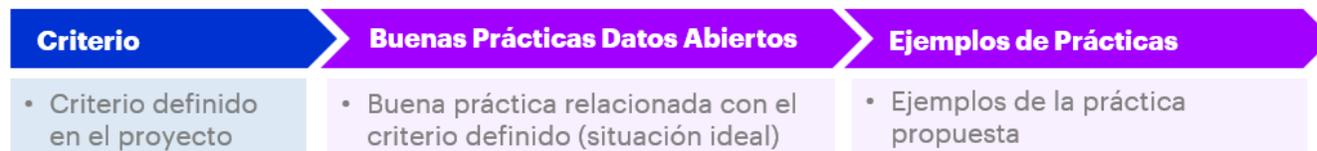
AGENDA FORMACIÓN



2. AGENDA FORMACIÓN

La presentación de la Guía Metodológica se estructura en tres grandes apartados: **Metadatos**, **Formatos reutilizables** y **Datos**.

Para cada uno de los apartados se identifican los **principales criterios** para la publicación de datos de calidad, identificados a lo largo del **proyecto** y a partir de la **experiencia** de **Accenture**. Para cada criterio se indicarán las **mejores prácticas** asociadas y **ejemplos**.



Adicionalmente, también se detallarán los **principales errores** detectados en el **Portal de Datos Abiertos**, junto con un **ejemplo** y la **buena práctica** asociada a este error.

Además, se incluye un bloque relativo a **otras guías externas** de interés para las fuentes de datos del portal de Madrid.

Por último, se reserva un bloque de **preguntas** y la realización de una **encuesta** por parte de los **participantes** para la valoración de la sesión formativa.

Extra: Breve formación básica de Python

10:00 - 10:15 – Introducción

10:15 – 10:45 – Criterios Metadatos

10:45 – 11:15 – Criterios Formatos Reutilizables

11:15 – 12:00 – Criterios Datos

12:00 – 12:30 - **DESCANSO**

12:30 – 12:50 – Errores Comunes

12:50 – 13:00 – Referencias externas

13:00 - 13:15 – Dudas/cierre

13:15 – 13:30 – Encuesta final de la formación

13:30 - 14:00 – Formación básica Python

MEJORES PRÁCTICAS

METADATOS



3. MEJORES PRÁCTICAS. METADATOS.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">El nombre del conjunto de datos es correcto y claro	<ol style="list-style-type: none">Mantener los nombres de los conjuntos de datos lo más cortos y concisos posible, evitando excesiva longitud o ambigüedad.	<ol style="list-style-type: none">El dataset "Puntos Limpios de Proximidad" tiene un nombre corto y conciso.



3. MEJORES PRÁCTICAS. METADATOS.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">Existe descripción correcta para el dataset	<ol style="list-style-type: none">Proporcionar una descripción detallada del conjunto de datos que incluya su contenido, fuente y propósito, para que los usuarios puedan evaluar su relevancia y comprender su contenido y contexto.Redactar la descripción del conjunto de datos en un lenguaje claro y accesible para el público en general.	<ol style="list-style-type: none">Buena práctica: “Aparcamientos municipales para residentes (PAR). Listas de espera”.Práctica a mejorar: “Taxi. Reservas de paradas en vía pública”



3. MEJORES PRÁCTICAS. METADATOS.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">El dataset tiene asignado un sector correctamente	<ol style="list-style-type: none">Clasificar el conjunto de datos en un sector o categoría apropiada para facilitar su búsqueda y comprensión.Utilizar la taxonomía de la Norma Técnica de Interoperabilidad para la categorización en sectores, de modo que estén consistentes y bien definidos.	<ol style="list-style-type: none">El dataset "Contratos menores" tiene asignado correctamente su sector.El dataset "Placas conmemorativas de Madrid" utiliza la taxonomía de la NTI (pág. 24).

3. MEJORES PRÁCTICAS. METADATOS.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">Existe un conjunto de palabras clave correcto para localizar los conjuntos de datos	<ol style="list-style-type: none">Incluir palabras clave relevantes (keywords) que describan el contenido y el tema del conjunto de datos, para mejorar su descubrimiento y clasificación.Utilizar el símbolo “;” como separador de palabras clave en lugar de otros caracteres para mantener la consistencia.Comprobar que las palabras clave no están encerradas entre comillas dobles.	<ol style="list-style-type: none">Dataset “Centros Municipales de Mayores. Datos sobre número y perfil de socios y socias”.



3. MEJORES PRÁCTICAS. METADATOS.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">Existe la fecha desde/hasta del dataset	<ol style="list-style-type: none">Incluir la fecha de inicio y fin de la cobertura temporal del conjunto de datos para que los usuarios comprendan el periodo cubierto.	<ol style="list-style-type: none">El dataset "Madrid Salud. Estadísticas centro de protección animal" tiene fecha desde y hasta.



3. MEJORES PRÁCTICAS. METADATOS.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">Frecuencia de actualización correcta	<ol style="list-style-type: none">Indicar claramente la frecuencia con la que se actualiza el conjunto de datos para que los usuarios estén informados sobre su vigencia.Garantizar que el conjunto de datos se actualiza según lo indicado y que los usuarios puedan acceder a las versiones más recientes.	<ol style="list-style-type: none">Buena práctica: “Denuncias y sanciones sobre limpieza urbana y zonas verdes”.Práctica a mejorar: “Mercados Municipales: locales comerciales, nombres comerciales y actividades”.



3. MEJORES PRÁCTICAS. METADATOS.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">Existe un responsable del conjunto de datos.	<ol style="list-style-type: none">Especificar claramente el nombre y la afiliación del responsable del conjunto de datos, ya sea una persona, una organización gubernamental, una empresa u otra entidad.Mantener la información del responsable del conjunto de datos actualizada para reflejar cualquier cambio en la responsabilidad o en la información de contacto.	<ol style="list-style-type: none">Buena práctica: "Accesibilidad y movilidad en aceras y calzadas. Año 2020" 



3. MEJORES PRÁCTICAS. METADATOS.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">Existe un responsable del conjunto de datos.	<ol style="list-style-type: none">3. Especificar claramente las responsabilidades del responsable del conjunto de datos, incluyendo la gestión de actualizaciones, la resolución de problemas técnicos y la atención a las consultas de los usuarios.	<ol style="list-style-type: none">1. Buena práctica: "Accesibilidad y movilidad en aceras y calzadas. Año 2020"

3. MEJORES PRÁCTICAS. METADATOS.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">Existe un documento de estructura correctamente estructurado	<ol style="list-style-type: none">Proporcionar documentación clara y completa que explique el significado y la interpretación de los códigos utilizados en los datos. Incluir recursos adicionales, como diccionarios de códigos o glosarios, para facilitar la comprensión de los códigos utilizados en los datos.Proporcionar un documento detallado que explique la estructura y el esquema del conjunto de datos, incluyendo la definición de las variables y sus tipos.	<ol style="list-style-type: none">Buena práctica: “Agencia para el Empleo. Perfiles de personas inscritas”.Este dataset tiene documento de estructura en el portal, como acabamos de ver.



3. MEJORES PRÁCTICAS. METADATOS.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">Existe un documento de estructura correctamente estructurado	<ol style="list-style-type: none">Nombrar el documento de estructura de manera coherente y siguiendo un formato estándar para facilitar su identificación y acceso.Verificar que las columnas en los conjuntos de datos coincidan con el archivo de estructura proporcionado. Esto asegurará que la información esté correctamente ubicada y facilitará la comprensión de los datos.	<ol style="list-style-type: none">El documento de estructura del ejemplo anterior tiene un nombre coherente. Otro ejemplo de buena práctica es el documento del dataset “Aforos de tráfico en la ciudad de Madrid permanentes”.Las columnas del fichero “juegos_areas_infantiles_2017.xlsx” del dataset “Mobiliario urbano. Juegos en Áreas Infantiles” coinciden con las descritas en el documento de estructura.



3. MEJORES PRÁCTICAS. METADATOS.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">Existe la licencia correcta del dataset	<ol style="list-style-type: none">Utilizar una licencia estandarizada y reconocible que sea consistente con otras licencias comúnmente utilizadas en conjuntos de datos similares, y especificarla claramente para que los usuarios conozcan los términos y condiciones de uso.	<ol style="list-style-type: none">El dataset "Pagos a terceros" tiene una licencia estandarizada.

MEJORES PRÁCTICAS

FORMATOS
REUTILIZABLES



3. MEJORES PRÁCTICAS. FORMATOS REUTILIZABLES.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">Fichero formalmente bien construido	<ol style="list-style-type: none">Evitar dejar columnas sin nombres o por ejemplo hacer CSVs en los que la información de una misma fila no esté partida en varias celdas y así conseguir mejores prácticas de estructuración y organización de los datos para garantizar su calidad y facilidad de uso.	<ol style="list-style-type: none">El fichero “inventario instalaciones fotovoltaicas 2022.xlsx” está bien construido.

3. MEJORES PRÁCTICAS. FORMATOS REUTILIZABLES.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">No existen filas o columnas en blanco	<ol style="list-style-type: none">Identificar y eliminar cualquier fila o columna que esté completamente en blanco o que contenga solo información no relevante o redundante.Utilizar herramientas automatizadas de limpieza y procesamiento de datos que puedan identificar y eliminar de manera eficiente filas o columnas en blanco.	<ol style="list-style-type: none">El fichero "RecogidaContenedoresAceiteUsado.xlsx" contiene algunos campos en blanco en sus filas y columnas. No representa el criterio, pero sirve a modo de ejemplo en la presentación. 

3. MEJORES PRÁCTICAS. FORMATOS REUTILIZABLES.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">No existen filas o columnas en blanco	<ol style="list-style-type: none">Realizar pruebas de calidad adicionales después de la limpieza de datos para garantizar que el conjunto de datos final cumpla con los estándares de calidad establecidos y sea útil y confiable para los usuarios	<ol style="list-style-type: none">El fichero "RecogidaContenedoresAceiteUsado.xlsx" contiene algunos campos en blanco en sus filas y columnas. No representa el criterio, pero sirve a modo de ejemplo en la presentación.

3. MEJORES PRÁCTICAS. FORMATOS REUTILIZABLES.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">Formato correcto del encabezado	<ol style="list-style-type: none">Hacer que el encabezado de las columnas esté formateado correctamente (evitar anidación), sea coherente con el resto de información y se ajuste a las expectativas y convenciones establecidas.Verificar que la primera fila del fichero sea utilizada exclusivamente para los encabezados de las columnas o que no esté en la segunda fila. Evitar la presencia de datos o cualquier otra información en la primera fila que no esté relacionada con los encabezados de las columnas.	<ol style="list-style-type: none">El fichero "Abonados 2022.xlsx" cumple con las tres buenas prácticas mencionadas.El fichero "Órdenes de ejecución 2020.xls" tiene encabezados anidados. Es un ejemplo de práctica a mejorar.

3. MEJORES PRÁCTICAS. FORMATOS REUTILIZABLES.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">Formato correcto del encabezado	<ol style="list-style-type: none">3. Utilizar nombres de columnas descriptivos y claros que sean comprensibles para los usuarios. Evitar el uso de abreviaturas o códigos crípticos en los nombres de las columnas y optar por terminología clara y concisa. Asegurarse de que los nombres de las columnas sigan un formato de texto universal que sea compatible con diferentes sistemas y aplicaciones	<ol style="list-style-type: none">1. El fichero "Abonados 2022.xlsx" cumple con las tres buenas prácticas mencionadas.2. El fichero "Órdenes de ejecución 2020.xls" tiene encabezados anidados. Es un ejemplo de práctica a mejorar.

3. MEJORES PRÁCTICAS. FORMATOS REUTILIZABLES.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">Existen filas o columnas de totales o subtotales	<ol style="list-style-type: none">Identificar si existen filas o columnas que contengan totales o subtotales en el conjunto de datos. Documentar claramente estas filas o columnas y considerar si es necesario incluirlas o excluirlas en el análisis posterior. (en principio no debería haber columnas totalizadoras).	<ol style="list-style-type: none">El fichero “Agua Regenerada Madrid 2022.xlsx” tiene columnas totalizadoras, cuando, en principio, no debería tener. Es un ejemplo de Práctica a mejorar.

3. MEJORES PRÁCTICAS. FORMATOS REUTILIZABLES.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">No incluir hojas vacías y preferentemente una hoja por fichero excel y preferentemente, una hoja por fichero excel	<ol style="list-style-type: none">Eliminar cualquier hoja vacía o sin datos en el conjunto de datos en formato de Excel, ya que csv solo se tiene una hoja. Asegurarse de que todas las hojas contengan información relevante y útil.También se debe incorporar como buena práctica, porque puede suponer un riesgo desde el punto de vista de protección de datos de incluir en esa hoja secundaria información no deseada por error.	<ol style="list-style-type: none">El fichero "Edificios declarados en ruina legal y física 2021.xls" tiene una hoja en blanco, cuando no debería tener ninguna. Es un ejemplo de Práctica a mejorar.

3. MEJORES PRÁCTICAS. FORMATOS REUTILIZABLES.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">El fichero debe contener una única tabla de datos por hoja	<ol style="list-style-type: none">Asegurarse de que el fichero de datos contenga solo una tabla con la información relevante. Evitar la presencia de múltiples tablas o conjuntos de datos en el mismo fichero, ya que esto puede generar confusión y dificultar el análisis posterior.	<ol style="list-style-type: none">Como ejemplo de Práctica a mejorar, el fichero "Órdenes de ejecución 2020.xls" tiene más de una tabla. Sólo debería tener una.Otro ejemplo de Práctica a mejorar es el fichero "tenis_mesa_2020.xlsx" tiene dos hojas de Excel, es decir, dos tablas, cuando debería tener una.

3. MEJORES PRÁCTICAS. FORMATOS REUTILIZABLES.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">Mismo orden y número de columnas en todas las filas, series temporales y formatos	<ol style="list-style-type: none">Verificar que no se hayan agregado o eliminado columnas en los nuevos conjuntos de datos, ya que esto podría afectar la consistencia y el análisis posterior.Mantener una estructura consistente en todas las series temporales y formatos relacionados, asegurando el mismo orden y número de columnas. Esto facilitará la comparación y el procesamiento de los datos.	<ol style="list-style-type: none">Los ficheros "pmed_ubicacion_04-2023.xlsx" y "pmed_ubicacion_05-2023.xlsx" del dataset "Puntos de medidas de tráfico" tienen el mismo orden y número de columnas.



3. MEJORES PRÁCTICAS. FORMATOS REUTILIZABLES.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">El nombre del fichero de datos es consistente con la serie anterior.	<ol style="list-style-type: none">Mantener nombres coherentes y descriptivos para los ficheros de datos, asegurándose de que sean consistentes con la serie anterior. Esto facilitará la identificación y el seguimiento de los datos a lo largo del tiempo.	<ol style="list-style-type: none">Los ficheros del dataset “Accidentes de tráfico de la Ciudad de Madrid”, relativos a 2024 y 2023 tienen nombres coherentes. Los de 2024 se llaman “2024_Accidentalidad.xxx” y los de 2023 “2023_Accidentalidad.xxx”.

3. MEJORES PRÁCTICAS. FORMATOS REUTILIZABLES.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">• Uso de “;” como campo separador de caracteres	<ol style="list-style-type: none">1. Utilizar el símbolo de punto y coma “;” como separador de campos en lugar de otros caracteres como comas o tabulaciones. Mantener una consistencia y verificar que el uso del punto y coma no genere conflictos o ambigüedades con los datos mismos, especialmente si los datos contienen caracteres especiales o texto que también pueda incluir punto y coma.	<ol style="list-style-type: none">1. El fichero “ContenedoresRopa.csv” tiene el símbolo “;” como separador de caracteres.

3. MEJORES PRÁCTICAS. FORMATOS REUTILIZABLES.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none"> El conjunto de datos está dividido en distintas distribuciones de manera que cada una de ellas sea... 	<ol style="list-style-type: none"> Si el conjunto de datos es extenso o complejo, considerar dividirlo en distribuciones más pequeñas y manejables. Proporcionar la... 	<ol style="list-style-type: none"> El fichero "1er trimestre.csv" del dataset "SER. Tiques de aparcamiento", cuyo tamaño es tan grande que no se puede procesar...

 1er trimestre		11/28/2022 3:08 PM	Microsoft Excel Com...	1,207,107 KB
---	---	--------------------	------------------------	--------------



3. MEJORES PRÁCTICAS. FORMATOS REUTILIZABLES.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">No existe demasiada anidación en los datos.	<ol style="list-style-type: none">1. Evaluar si la anidación excesiva de los datos dificulta su análisis y comprensión con herramientas de análisis de datos (Excel, Python, etc.). Considerar si es necesario simplificar la estructura de los datos para facilitar su uso.	<ol style="list-style-type: none">1. El fichero "descuentos202404.json" apenas tiene anidación.2. Ejemplo fichero json en formato Excel para ver anidación.3. Un ejemplo de fichero con demasiada anidación es "ocupacion_rodajes_2023.xml". Esto es una Práctica a mejorar.

3. MEJORES PRÁCTICAS. FORMATOS REUTILIZABLES.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">Codificación correcta de caracteres	<ol style="list-style-type: none">Utilizar una codificación estándar y ampliamente aceptada, como UTF-8, para garantizar la compatibilidad	<ol style="list-style-type: none">El fichero "Actuaciones_UDC.csv" tiene codificación utf-8.
<pre>modulos_2022.csv x C:\Users> riglesias > Downloads > modulos_2022.csv 1 Módulo; Nombre del Cliente; Cif/Nif; Ubicación; Teléfono; Página WEB; Código Área Actividad; Descripción Área Actividad 2 AB1 ; ALJOFER, S.A. 3 AB1 ; HERMANOS FRANCH TUTELAR, S.A. 4 AB1 ; JOSE FRANCH S.L. 5 AB1 ; FRUTAS Y VERDURAS J&R BARGUEÑO, S.L. 6 AB3 ; ALJOFER, S.A. 7 AB3 ; HERMANOS FRANCH TUTELAR, S.A. 8 AB3 ; JOSE FRANCH S.L. 9 AB3 ; FRUTAS Y VERDURAS J&R BARGUEÑO, S.L. 10 AB4 ; LORENZO IZQUIERDO, S.L. 11 AB5 ; ALJOFER, S.A.</pre>		

3. MEJORES PRÁCTICAS. FORMATOS REUTILIZABLES.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">Organización vertical de la información, en vez de horizontal.	<ol style="list-style-type: none">Agrupar los datos en categorías principales que representen conceptos más generales o de alto nivel, y luego desglosar estos conceptos en subcategorías más específicas o detalladas.	<ol style="list-style-type: none">El fichero “Agua Regenerada Madrid 2022.xlsx” tiene una organización horizontal. Es un ejemplo de práctica a mejorar.El fichero “Abonados 2022.xlsx” del dataset “Deportes. Abonados en Centros Deportivos Municipales” tiene una organización vertical. Es un ejemplo de buena práctica.

3. MEJORES PRÁCTICAS. FORMATOS REUTILIZABLES.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">Identificación en los datos, del año y mes a que hacen referencia.	<ol style="list-style-type: none">El uso de un formato estándar de fecha, preferiblemente YYYY-MM (año-mes).Actualizar regularmente la fecha de referencia para reflejar la información más reciente.	<ol style="list-style-type: none">1. El fichero “ActuacionesBomberos_2024.xlsx” tiene cada registro etiquetado con la fecha, y la fecha de referencia está incluida en el nombre del fichero.

3. MEJORES PRÁCTICAS. FORMATOS REUTILIZABLES.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">Mes mejor en formato numérico en vez de texto, para permitir una ordenación cronológica de los meses, en vez de alfabética.	<ol style="list-style-type: none">Mantener la consistencia en la representación numérica de los meses en todos los conjuntos de datos relacionados.Se recomienda validar la precisión de los valores numéricos asignados a los meses para evitar errores y garantizar una ordenación cronológica precisa y confiable de los datos.	<ol style="list-style-type: none">2. El fichero "reservas_moto.xls" del dataset "Moto. Reservas Moto" tiene la columna de fechas como valor de fecha.

3. MEJORES PRÁCTICAS. FORMATOS REUTILIZABLES.

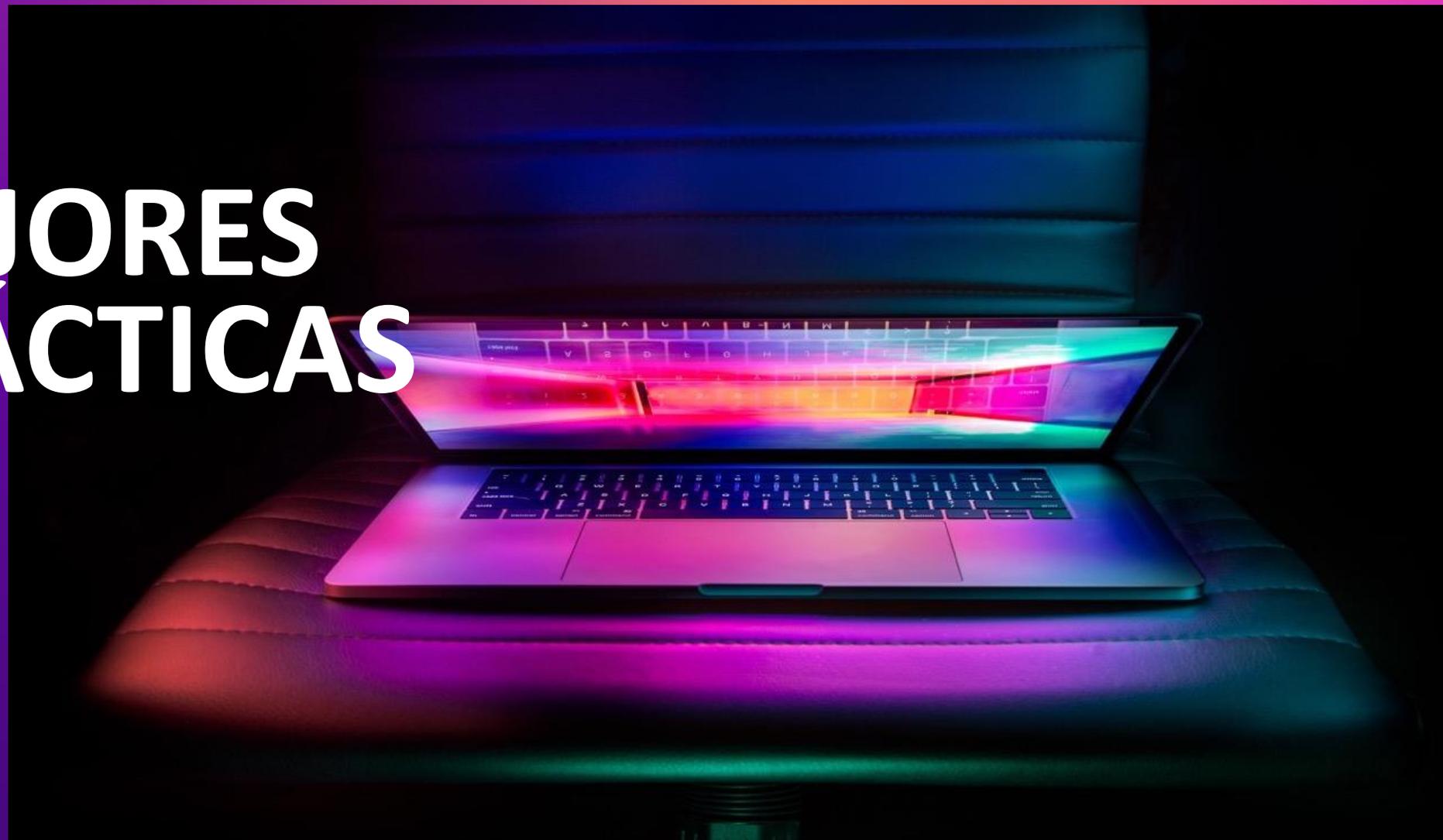
Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">Utilizar el mayor número de formatos posibles, para que los datos sean más accesibles.	<ol style="list-style-type: none">Proporcionar los datos en múltiples formatos, incluyendo CSV, JSON, XML, RDF, entre otros, para adaptarse a las diferentes necesidades y preferencias de los usuarios.	<ol style="list-style-type: none">El dataset "Actividades Culturales y de Ocio Municipal en los próximos 100 días" tiene varios formatos para descargar.El fichero "Defunciones clasificadas por nacionalidad y sexo según Distrito y Barrio. 1 Enero 2022" es un ejemplo de mala práctica.

3. MEJORES PRÁCTICAS. FORMATOS REUTILIZABLES.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">No existen metadatos de autor	<ol style="list-style-type: none">Revisión y limpieza de metadatos: Antes de publicar cualquier conjunto de datos, realizar un proceso de revisión para identificar y eliminar cualquier metadato de autor.	<ol style="list-style-type: none">El dataset “Aparcamientos municipales para residentes (PAR). Valor de las plazas - Portal de datos abiertos del Ayuntamiento de Madrid” no tiene metadatos de autor para el fichero de 2024, mientras que si tiene en el de 2023.

MEJORES PRÁCTICAS

DATOS



3. MEJORES PRÁCTICAS. DATOS.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">Orden lógico de las columnas	<ol style="list-style-type: none">Mantener siempre el mismo orden de las columnas en todos los conjuntos de datos similares. Esto facilita la automatización y el análisis por parte de los usuarios.Colocar primero las columnas que contienen identificadores únicos y datos temporales. Esto permite una rápida referencia y seguimiento de cada registro.	<ol style="list-style-type: none">El dataset “Tráficos. Cabezas de semáforo” tiene los campos ordenados de manera lógica.



3. MEJORES PRÁCTICAS. DATOS.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">Orden lógico de las columnas	<ol style="list-style-type: none">3. Ordenar las columnas siguiendo una estructura jerárquica, desde divisiones administrativas más amplias (distrito) hasta las más específicas (barrio y dirección).4. Adoptar estándares reconocidos para la organización de datos. Por ejemplo, utiliza formatos y convenciones comúnmente aceptados para fechas, coordenadas y otros datos específicos.	<ol style="list-style-type: none">1. El dataset “Tráficos. Cabezas de semáforo” tiene los campos ordenados de manera lógica.



3. MEJORES PRÁCTICAS. DATOS.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">Valores de distribución de cada columna coherentes con la serie anterior	<ol style="list-style-type: none">Realizar una verificación de consistencia en los valores de las columnas, asegurándose de que estén dentro del rango esperado y sean coherentes con la serie anterior. Esto ayudará a identificar posibles errores o discrepancias en los datos.Verificar si hay cambios drásticos en los valores de la media y desviación estándar y validar si la media se encuentra dentro de un rango esperado y coherente con la serie anterior.	<ol style="list-style-type: none">Los valores entre columnas son consistentes si se compara la serie de abril con la de mayo en el dataset "Presupuestos. Ejecución mensual ejercicio 2023". La media mantiene su tendencia. Lo mismo ocurre con la desviación estándar "std".



3. MEJORES PRÁCTICAS. DATOS.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">Valores de distribución de cada columna coherentes con la serie anterior	<ol style="list-style-type: none">Verificar que los valores máximos en cada columna no excedan los límites establecidos en la serie anterior. Esto ayudará a identificar valores atípicos o errores en los datos.Verificar que los valores mínimos en cada columna no sean inferiores a los límites establecidos en la serie anterior. Esto ayudará a identificar posibles errores o discrepancias en los datos.	<ol style="list-style-type: none">Analizando los mismos ficheros del ejemplo previo, los valores máximos y mínimos no exceden la serie anterior. 



3. MEJORES PRÁCTICAS. DATOS.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">Valores de distribución de cada columna coherentes con la serie anterior	<p>6. Analizar cualquier cambio drástico en la frecuencia del valor más común y determinar si hay razones válidas para ese cambio. Verificar si la frecuencia del valor más común sigue siendo consistente con la serie anterior.</p>	<p>4. En el dataset "Convenios", el valor más frecuente disminuye en un 21% de 2022 a 2023. Tal vez resulte de interés analizar por qué ocurre esto.</p> 



3. MEJORES PRÁCTICAS. DATOS.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">Valores de distribución de cada columna coherentes con la serie anterior	<p>7. Comprobar si los tipos de frecuencia asignados (+ para las más frecuentes y - para las menos frecuentes) se corresponden con las expectativas y la serie anterior. Validar si la asignación de frecuencias es coherente y sigue una lógica adecuada en función de los datos y las tendencias observadas.</p>	<p>5. Para el fichero anterior, se comprueba que los tipos de frecuencia asignados se mantienen de una serie a otra.</p>

3. MEJORES PRÁCTICAS. DATOS.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">Los tipos de campos se ajustan a lo esperado	<ol style="list-style-type: none">Verificar que los tipos de datos utilizados en los campos sean apropiados y se ajusten a la naturaleza de la información. Esto evitará problemas de interpretación y permitirá un análisis preciso.	<ol style="list-style-type: none">En el fichero “plan operativo gobierno 02062 023.xls” los tipos de datos se ajustan a lo esperado según los nombres de las columnas.

3. MEJORES PRÁCTICAS. DATOS.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">Asignación de un ID único	<ol style="list-style-type: none">Asignar un identificador único a cada registro o entidad en los datos (columna) para facilitar su identificación y seguimiento. Utilizar un esquema de ID claro y consistente para evitar duplicados o confusiones.	<ol style="list-style-type: none">El fichero “servicios_sociales_personasatendidas_2023.xlsx” tiene un identificador único. La columna “Secuencia”.

3. MEJORES PRÁCTICAS. DATOS.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">Los valores de datos de tipo fecha y fecha/hora deben describirse en formato ISO 8601	<ol style="list-style-type: none">Representar las fechas y horas en formato ISO 8601 para facilitar la interoperabilidad y la comprensión de los datos temporales. Evitar ambigüedades y asegurarse de que las fechas se interpreten correctamente en diferentes sistemas y aplicaciones	<ol style="list-style-type: none">Como ejemplo de Práctica a mejorar en el uso del formato ISO 8601, que establece que las fechas completas deben representarse: YYYY-MM-DD (Año-Mes-Día), se tiene el fichero "TAXI Flota Diario.xls", el cual no tiene la columna "Fecha inicio de prestación del servicio de taxi" de acuerdo a este formato. Por ejemplo, la primera fila debería ser "2021/06/18".



3. MEJORES PRÁCTICAS. DATOS.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">Cumplimiento de codificación para información de barrios y distritos	<ol style="list-style-type: none">Asegurarse de que la información relacionada con los distritos y barrios cumpla con las pautas y estándares de codificación establecidos. Esto garantizará la coherencia y la interoperabilidad de los datos. Utilizar el estándar definido en el portal: Barrios municipales de Madrid y Distritos municipales de Madrid	<ol style="list-style-type: none">Los dataset “Barrios municipales de Madrid” y “Distritos municipales de Madrid” tienen los códigos de distrito y barrios de acuerdo con el estándar.Verificar con el fichero “poblacion_1_enero.csv”. Vemos que sigue la codificación estándar.



3. MEJORES PRÁCTICAS. DATOS.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">Formato de dirección válida	<ol style="list-style-type: none">Establecer un formato estándar para la representación de direcciones dentro del conjunto de datos. Esto puede incluir el orden de los campos (tipo de vial, nombre de la vía, número de edificio), el uso de separadores (como comas o espacios) y la capitalización de las palabras.	<ol style="list-style-type: none">En las estructuras de datos de los datasets, ofrecer una guía en los campos relacionados a direcciones. Un ejemplo es la ficha de estructura del dataset “Mobiliario Urbano. Papeleras”.



3. MEJORES PRÁCTICAS. DATOS.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">• Formato de dirección válida	<ol style="list-style-type: none">2. Aprovechar al máximo los campos disponibles en el conjunto de datos para incluir toda la información relevante para una dirección completa y precisa. Esto puede incluir no solo el tipo de vial, el nombre de la vía y el número de edificio, sino también campos adicionales como el código postal, el distrito o la ciudad.	<ol style="list-style-type: none">1. En las estructuras de datos de los datasets, ofrecer una guía en los campos relacionados a direcciones. Un ejemplo es la ficha de estructura del dataset "Mobiliario Urbano. Papeleras".

3. MEJORES PRÁCTICAS. DATOS.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">Las coordenadas se expresan en latitud y longitud y en el formato correcto	<ol style="list-style-type: none">Utilizar el sistema de coordenadas ETRS89 y asegurarse de que las longitudes y latitudes estén representadas en formato decimal. Esto garantizará la consistencia y facilitará el análisis geoespacial.	<ol style="list-style-type: none">El fichero "ContenedoresRopa.xlsx" no tiene ningún valor fuera de rangoEl rango de valores normal para Madrid ronda el -3,xxx y el 40,xxx en longitud y latitud, respectivamente. Este rango de valores viene dado por https://www.ign.es/web/coordenadas-de-estaciones-ergnss. Este es un ejemplo de buena práctica.



3. MEJORES PRÁCTICAS. DATOS.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">Los valores nulos y no nulos se ajustan a lo esperado	<ol style="list-style-type: none">Verificar que los valores no nulos sean consistentes con las expectativas y requisitos de los datos. Esto ayudará a garantizar la integridad y calidad de los datos. <p>Es importante diferenciar entre valor nulo y valor cero. Un valor cero es un valor informado, mientras que un valor nulo indica ausencia de información.</p>	<ol style="list-style-type: none">Mientras que el fichero "mupis_pilas.xlsx" se ajusta a lo esperado, todos los campos están correctamente informados, en el "datos_de_gestion.xls", del dataset "Multas de circulación" hay mayor número de celdas vacías. Es un ejemplo de práctica a mejorar



3. MEJORES PRÁCTICAS. DATOS.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">Los valores nulos y no nulos se ajustan a lo esperado	<ol style="list-style-type: none">Comprobar si hay una disminución significativa en el número de celdas no vacías y verificar si el número de celdas no vacías sigue siendo coherente con la serie anterior y si cumple con los requisitos esperados.	<ol style="list-style-type: none">En el dataset "Multas de circulación" hay un mayor número de celdas vacías. Este sería un ejemplo en el que habría que analizar el porqué de esta disminución, y si los valores siguen siendo coherentes con la serie anterior.



3. MEJORES PRÁCTICAS. DATOS.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">• X e Y correctamente representadas	<ol style="list-style-type: none">1. Verificar que las coordenadas X e Y estén correctamente representadas y se ajusten al sistema de coordenadas esperado. Esto asegurará la precisión y la coherencia espacial de los datos.	<ol style="list-style-type: none">1. El rango de valores orientativos para la Ciudad de Madrid para la X es de 433000 a 391000 en metros, y el de Y es de 4960000 a 4910000 en metros. Si no estuviese entre esos rangos, probablemente estaría fuera del término de Madrid. Este rango de valores viene dado por https://www.ign.es/web/coordenadas-de-estaciones-ergnss. Cualquier fichero con coordenadas X e Y debe tener valores entre esos rangos.

3. MEJORES PRÁCTICAS. DATOS.

Criterio	Buenas Prácticas	Ejemplos					
<ul style="list-style-type: none">Decimales representados con una coma	<ol style="list-style-type: none">Utilizar la coma como separador decimal en lugar de puntos u otros caracteres. Mantener la consistencia en la representación de los valores decimales en todo el conjunto de datos.	<ol style="list-style-type: none">La columna del fichero de 2015 “modificaciones inscritas 2015.xls” del dataset “Actividad Contractual” es correcta. Hay un ejemplo marcado en naranja de cómo no se debe utilizar el punto como separador decimal: <table border="1" data-bbox="1727 943 1989 1232"><thead><tr><th>IMPORTE DE LA MODIFICACIÓN</th></tr></thead><tbody><tr><td>0.00 €</td></tr><tr><td>57,957.14 €</td></tr><tr><td>6229,67</td></tr><tr><td>-67,115.43 €</td></tr></tbody></table>	IMPORTE DE LA MODIFICACIÓN	0.00 €	57,957.14 €	6229,67	-67,115.43 €
IMPORTE DE LA MODIFICACIÓN							
0.00 €							
57,957.14 €							
6229,67							
-67,115.43 €							



3. MEJORES PRÁCTICAS. DATOS.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">No se deben utilizar caracteres de formato de "miles".	<ol style="list-style-type: none">Evitar el uso de caracteres como puntos o comas para separar miles en números. Mantener la coherencia en el formato numérico utilizado en todo el conjunto de datos.	<ol style="list-style-type: none">El fichero "centroproteccionanimal.xls" no utiliza puntos ni comas para separar miles. El campo ratios colonias utiliza el punto como separador decimal. Este es un ejemplo de buena práctica.

3. MEJORES PRÁCTICAS. DATOS.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">No se deben incluir ceros a la izquierda.	<ol style="list-style-type: none">Eliminar cualquier cero no significativo a la izquierda de los números. Asegurarse de que los números se representen de manera consistente y sigan un formato numérico estándar.	<ol style="list-style-type: none">Un ejemplo de Práctica a mejorar, teniendo ceros a la izquierda, sería "0198". En el caso de que esa celda fuese formato texto, no sería lo mismo que si se tiene "198".

3. MEJORES PRÁCTICAS. DATOS.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">Las unidades de medida y monedas deben indicarse por separado, o en el nombre de las columnas.	<ol style="list-style-type: none">Especificar claramente las unidades de medida y las monedas utilizadas en los datos. Evitar ambigüedades proporcionando etiquetas o metadatos claros para las columnas que contienen información de unidades o monedas.	<ol style="list-style-type: none">El fichero “EstadoParquesHistoricoSingulares Forestales2017.xlsx” del dataset “Arbolado en parques y zonas verdes de Madrid” tiene correctamente especificadas las unidades en su documento de estructura.

3. MEJORES PRÁCTICAS. DATOS.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">Valores de distribución de cada columna coherentes con la serie anterior	<ol style="list-style-type: none">Verificar que el número de filas en los nuevos conjuntos de datos sea aproximadamente igual o similar al número de filas en la serie anterior. Esto garantizará que no se haya perdido o agregado información incorrectamente.	<ol style="list-style-type: none">El número de filas entre series en el dataset "Aparcamientos (PAR) - valor de las plazas" es similar. 



3. MEJORES PRÁCTICAS. DATOS.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">Confidencialidad y anonimización de los datos	<ol style="list-style-type: none">Garantizar la confidencialidad y privacidad de los datos sensibles. Aplicar técnicas adecuadas de anonimización de datos para proteger la identidad de las personas o entidades involucradas. No se puede publicar nada si no hay una norma o ley que lo habilite. Tampoco se pueden publicar campos que puedan identificar a una persona.	Depende del caso



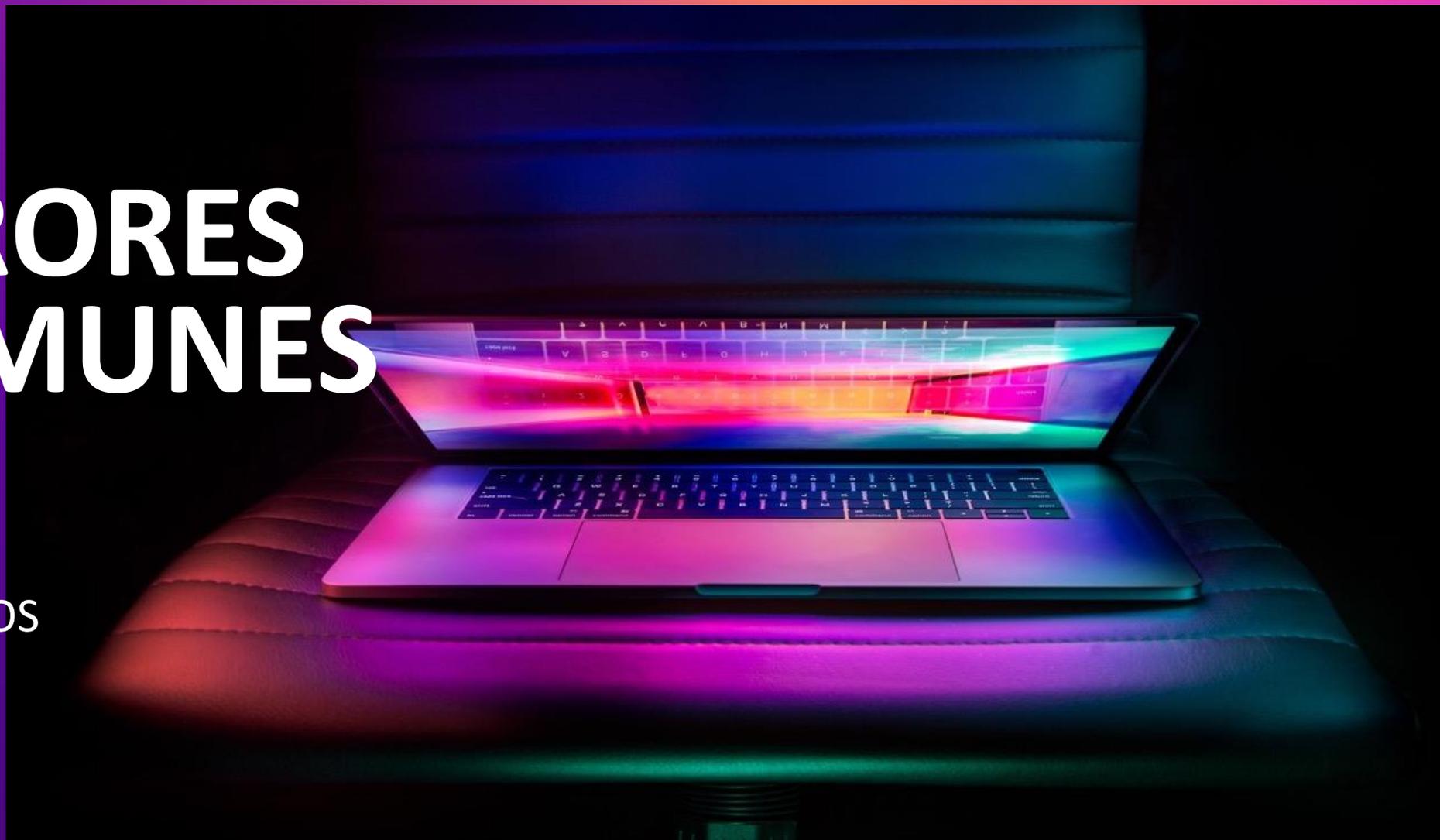
3. MEJORES PRÁCTICAS. DATOS.

Criterio	Buenas Prácticas	Ejemplos
<ul style="list-style-type: none">Consistencia entre datos del portal de Madrid y fuentes externas (Madrid.es y Banco de Datos de Estadística)	<ol style="list-style-type: none">Asegurar la coherencia y consistencia entre los datos del portal del Ayuntamiento de Madrid y fuentes externas (como Madrid.es y el Banco de Datos de Estadística del Ayto. de Madrid). Identificar cualquier discrepancia o inconsistencia entre los conjuntos de datos y tomar medidas para corregirlas o aclararlas.	<ol style="list-style-type: none">Los datos de 2023 del dataset “Inspecciones de prevención de incendios” entre los portales Madrid.es y el portal del ayuntamiento de Madrid son consistentes y coherentes. Se pueden consultar en los links del Portal del Ayuntamiento de Madrid y Madrid.es



ERRORES COMUNES

METADATOS



3. ERRORES COMUNES. METADATOS

Error típico	Impacto y evolución	Ejemplo	Vinculación con buena práctica
<ul style="list-style-type: none">Descripción demasiado corta	<ol style="list-style-type: none">Se verifica en la actividad 1 de la fase 3. De los 577 ficheros analizados, 110 presentaban este error, es decir, el 19%.En la fase 1 casi el 37% de los ficheros presentaban este error (200 ficheros de 541). Se ha mejorado un 18% desde el inicio del contrato.	<ol style="list-style-type: none">Ejemplo de Práctica a mejorar: "Taxi. Reservas de paradas en vía pública".Ejemplo de buena práctica: "Aparcamientos municipales para residentes (PAR). Listas de espera".	<ol style="list-style-type: none">Se vincula directamente con la buena práctica "Proporcionar una descripción detallada del conjunto de datos que incluya su contenido, fuente y propósito, para que los usuarios puedan evaluar su relevancia.", del criterio 3.1.2 de metadatos.

3. ERRORES COMUNES. METADATOS

Error típico	Impacto y evolución	Ejemplo	Vinculación con buena práctica
<ul style="list-style-type: none">No existencia documento de estructura	<ol style="list-style-type: none">Se verifica en la actividad 1 de la fase 3. De los 577 ficheros analizados, 273 presentaban este error, es decir, el 47%.En la fase 1 casi el 60% (320 ficheros de 541) de los ficheros presentaban este error. Se ha mejorado un 13% desde el inicio del contrato.	<ol style="list-style-type: none">El dataset "Encuesta de satisfacción de participantes en el Programa Marcha Nórdica" no tiene documento de estructura en el portal.	<ol style="list-style-type: none">Se vincula directamente con la buena práctica "Proporcionar un documento detallado que explique la estructura y el esquema del conjunto de datos, incluyendo la definición de las variables y sus tipos.", del criterio 3.1.8 de metadatos.

ERRORES COMUNES

FORMATOS
REUTILIZABLES



3. ERRORES COMUNES. FORMATOS REUTILIZABLES.

Error típico	Impacto	Ejemplo	Vinculación con buena práctica
<ul style="list-style-type: none">Codificación diferente a UTF-8	<ol style="list-style-type: none">Se verifica en la actividad 3 de la fase 3 a nivel de fichero, no de dataset como en los puntos anteriores. El promedio de ficheros con este error es 51.11%.En la fase 2 el promedio de ficheros con este error era aproximadamente del 50%. Se mantiene la tendencia en la fase 3.	<ol style="list-style-type: none">En el ejemplo del criterio 3.2.13 se observa qué ocurre cuando se tiene un fichero no codificado en utf-8.	<ol style="list-style-type: none">Se vincula directamente con la buena práctica “Utilizar una codificación estándar y ampliamente aceptada, como UTF-8, para garantizar la compatibilidad y evitar problemas de interpretación de caracteres especiales.”, del criterio 3.2.13.



3. ERRORES COMUNES. FORMATOS REUTILIZABLES.

Error típico	Impacto	Ejemplo	Vinculación con buena práctica
<ul style="list-style-type: none">Demasiada anidación en los ficheros JSON/XML	<ol style="list-style-type: none">Se verifica en la actividad 3 de la fase 3 a nivel de fichero. El promedio de ficheros con este error es 2.61%. Apenas existen ficheros con demasiada anidación.En la fase 2 el promedio de ficheros con este error era un 1.14%. Aumenta un poquito en la fase 3.	<ol style="list-style-type: none">En los ejemplos del criterio 3.2.12 se observan dos casos, uno en el que apenas existe anidación, y otro en el que la anidación es excesiva.	<ol style="list-style-type: none">Se vincula directamente con la buena práctica "Evaluar si la anidación excesiva de los datos dificulta su análisis y comprensión con herramientas de análisis de datos (Excel, Python, etc.). Considerar si es necesario simplificar la estructura de los datos para facilitar su uso.", del criterio 3.2.12.

3. ERRORES COMUNES. FORMATOS REUTILIZABLES.

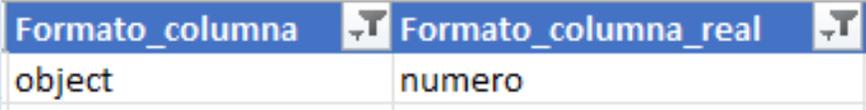
Error típico	Impacto	Ejemplo	Vinculación con buena práctica
<ul style="list-style-type: none">• CSV con “,” como separador de caracteres	<ol style="list-style-type: none">1. Verificado en múltiples actividades a lo largo de las fases del proyecto.	<ol style="list-style-type: none">1. El fichero “APARCABICIS_2023.csv” tiene como separador de caracteres el símbolo “,”.	<ol style="list-style-type: none">1. Se vincula directamente con la buena práctica “Utilizar el símbolo de punto y coma “;” como separador de campos en lugar de otros caracteres como comas o tabulaciones. Mantener una consistencia y verificar que el uso del punto y coma no genere conflictos o ambigüedades con los datos mismos, especialmente si los datos contienen caracteres especiales o texto que también pueda incluir punto y coma.”, del criterio 3.2.10.

ERRORES COMUNES

DATOS



3. ERRORES COMUNES. DATOS.

Error típico	Impacto	Ejemplo	Vinculación con buena práctica
<ul style="list-style-type: none">Muchos campos no se  pero están como texto).	<ol style="list-style-type: none">Hay 231 ficheros de 430 de 510, el 43.49%. Depende de las columnas de cada dataset.En la fase 2 la media de los 12 packs son 42.52%. El promedio de la fase 3 mencionado arriba es de un 48.60%. Se ha empeorado un 6%.	<ol style="list-style-type: none">Para ver un ejemplo de tipos de datos erróneos, vaya al ejemplo del apartado 3.2.2.Otro ejemplo: Imaginemos que los números del ejemplo tienen formato texto. Entonces no se podrán realizar operaciones con ellos.	<ol style="list-style-type: none">Se vincula directamente con la buena práctica “Verificar que los tipos de datos utilizados en los campos sean apropiados y se ajusten a la naturaleza de la información. Esto evitará problemas de interpretación y permitirá un análisis preciso.”, del criterio 3.2.2.

3. ERRORES COMUNES. DATOS.

Error típico	Impacto	Ejemplo	Vinculación con buena práctica
<ul style="list-style-type: none">Falta de asignación de un ID único a cada fila del fichero de datos	<ol style="list-style-type: none">Se verifica en la actividad 3 de la fase 3 a nivel de fichero. El promedio de ficheros con este error es 79.41%.En la fase 2 el promedio de ficheros con este error es 73.57%. El promedio de la fase 3 mencionado arriba es de un 79.41%. Se ha aumentado un 6%.	<ol style="list-style-type: none">El fichero "reconocimientos medicos basicos 2021.xlsx" no tiene ID único. Se observa su ausencia en las primeras columnas. <p>Para ver un fichero con ID único, vaya al ejemplo del apartado 3.3.3.</p>	<ol style="list-style-type: none">Se vincula directamente con la buena práctica "Asignar un identificador único a cada registro o entidad en los datos (columna) para facilitar su identificación y seguimiento. Utilizar un esquema de ID claro y consistente para evitar duplicados o confusiones.", del criterio 3.3.3.

3. ERRORES COMUNES. DATOS.

Error típico	Impacto	Ejemplo	Vinculación con buena práctica
<ul style="list-style-type: none">Decimales no representados por coma	<ol style="list-style-type: none">Se verifica en la actividad 3 de la fase 3 a nivel de fichero. El promedio de ficheros con este error es un 28%.En la fase 2 el promedio de ficheros con este error es del 43.45%. El promedio de la fase 3 mencionado arriba es de un 28%. Se ha mejorado un 15%.	<ol style="list-style-type: none">Se puede observar un ejemplo de práctica a mejorar en el ejemplo del criterio 3.3.10. Aparece una celda con separador distinto de “.”.	<ol style="list-style-type: none">Se vincula directamente con la buena práctica “Utilizar la coma como separador decimal en lugar de puntos u otros caracteres. Mantener la consistencia en la representación de los valores decimales en todo el conjunto de datos.”, del criterio 3.3.10.

3. ERRORES COMUNES. DATOS.

Error típico	Impacto	Ejemplo	Vinculación con buena práctica
<ul style="list-style-type: none">Fecha en formato no compatible con ISO 8601	<ol style="list-style-type: none">Se verifica en la actividad 3 de la fase 3 a nivel de fichero. El promedio de ficheros con este error es 21.91%.En la fase 2 el promedio de ficheros con este error es del 74.70%. El promedio de la fase 3 mencionado arriba es de un 21.91%. Se ha mejorado un 52.79%	<ol style="list-style-type: none">Se puede observar un ejemplo de práctica a mejorar en el ejemplo del criterio 3.3.4. El formato no corresponde con el requerido: YYYY-MM-DD.	<ol style="list-style-type: none">Se vincula directamente con la buena práctica "Representar las fechas y horas en formato ISO 8601 para facilitar la interoperabilidad y la comprensión de los datos temporales. Evitar ambigüedades y asegurarse de que las fechas se interpreten correctamente en diferentes sistemas y aplicaciones.", del criterio 3.3.4.

GUÍAS EXTERNAS



5. GUÍAS EXTERNAS - INICIATIVA CIUDADES ABIERTAS

La iniciativa *Ciudades Abiertas* es un convenio de ayuntamientos en España para la colaboración, reutilización e interoperabilidad de portales públicos de datos abiertos.

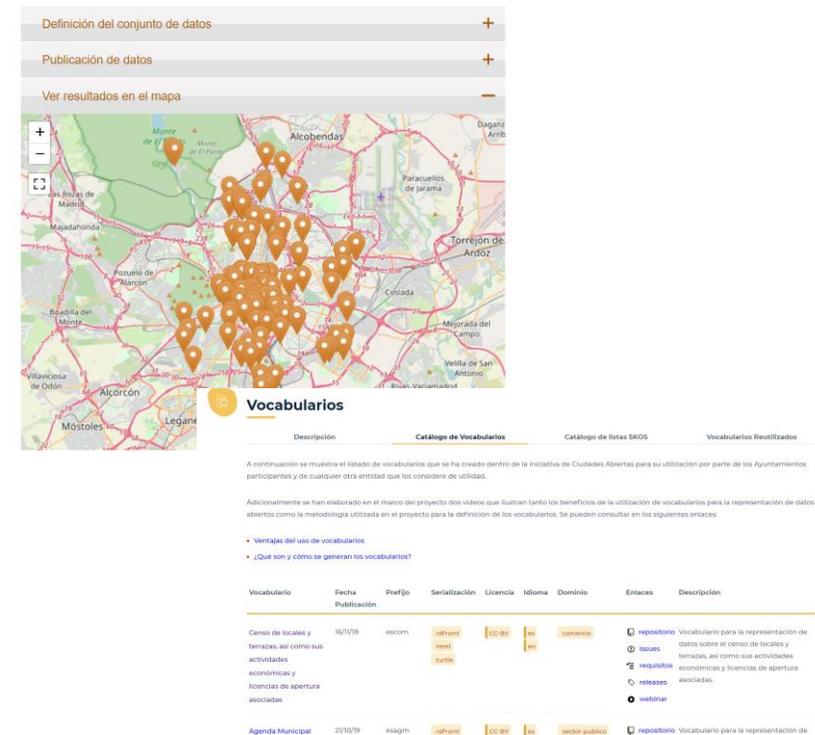
Principales criterios de calidad indicados en el checklist de la guía de la iniciativa de Ciudades Abiertas:

FORMATOS:

- En el caso de representar polígonos u otras formas geométricas, se recomienda usar el formato WKT (ejemplo dataset [“Actividades Culturales y de Ocio Municipal en los próximos 100 días”](#)).

DATOS:

- Siempre que sea posible, utilizar valores en los campos que estén estandarizados (ejemplo vocabularios [“https://ciudades-abiertas.es/vocabularios/”](https://ciudades-abiertas.es/vocabularios/)).
- Si no se tiene información completa sobre una fecha, no debe representarse como tal, sino en diversas columnas



The screenshot displays the 'Ciudades Abiertas' interface. At the top, there are navigation tabs: 'Definición del conjunto de datos', 'Publicación de datos', and 'Ver resultados en el mapa'. Below these is a map of Madrid with numerous orange location pins. A 'Vocabularios' section is visible, featuring a table with columns for 'Vocabulario', 'Fecha Publicación', 'Prefijo', 'Serialización', 'Licencia', 'Idioma', 'Dominio', 'Enlaces', and 'Descripción'. The table lists various datasets such as 'Censo de locales y terrazas' and 'Agenda Municipal'.

Vocabulario	Fecha Publicación	Prefijo	Serialización	Licencia	Idioma	Dominio	Enlaces	Descripción
Censo de locales y terrazas, así como sus actividades económicas y licencias de apertura asociadas	16/1/18	escom	shp/xml html turtle	CC BY	es	comercio	repositorio issues repositorio retos webinar	Vocabulario para la representación de datos sobre el censo de locales y terrazas, así como sus actividades económicas y licencias de apertura asociadas.
Agenda Municipal	2/10/19	esagm	shp/xml	CC BY	es	sector-publico	repositorio	Vocabulario para la representación de datos de la agenda municipal.

5. GUÍAS EXTERNAS - INICIATIVA APORTA

La *Iniciativa Aporta (datos.gob)* es una iniciativa promovida por el Gobierno de España con el objetivo de fomentar la **apertura y reutilización** de la información del **sector público**.

Principales criterios de calidad interesantes indicados en el checklist de la guía de la iniciativa Aporta (datos.gob):

METADATOS:

- Utilizar vocabularios controlados siempre que sea posible.

FORMATOS:

- Evitar formatos de datos no procesables.

DATOS:

- Estandarizar valores de datos.
- Evitar la mezcla de rangos en un mismo conjunto de datos

```
<dct:accrualPeriodicity>
  <dct:Frequency rdf:about="http://dataset/Frequency">
    <rdf:value>
      <time:DurationDescription rdf:about="http://dataset/DurationDescription">
        <time:years rdf:datatype="http://www.w3.org/2001/XMLSchema#decimal"1</time:years>
      </time:DurationDescription>
    </rdf:value>
  </dct:Frequency>
</dct:accrualPeriodicity>
```



```
<dct:accrualPeriodicity>
  <dct:Frequency rdf:about="http://publications.europa.eu/resource/authority/frequency/DAILY"/>
</dct:accrualPeriodicity>
```



marca	año	consumo	ventas	potencia	aceleración
ford torino	1970	Alto	2.50	De 100 a 150	12
buick skylark 320	1970	Medio	2.63	De 150 a 200	11.5
plymouth satellite	1970	Medio	2.37	De 150 a 200	11
chvrolet chevelle malibu	1970	Bajo	2.40	Mas de 200	13



marca	año	consumo	ventas	potencia	aceleración
ford torino	1970	Alto	2.50	130	12
buick skylark 320	1970	Medio	2.63	165	11.5
plymouth satellite	1970	Medio	2.37	150	11
chvrolet chevelle malibu	1970	Bajo	2.40	210	13



5. GUÍAS EXTERNAS - INICIATIVA APORTA

A continuación, se puede visualizar un *infográfico* con los *principales criterios (pautas)* para garantizar la *calidad de los datos abiertos*, indicado por la guía de datos.gov:

datos.gov.es

PAUTAS GENERALES PARA GARANTIZAR LA CALIDAD DE LOS DATOS ABIERTOS

EVITAR FORMATOS DE DATOS NO PROCESABLES
Cualquier informe elaborado a partir de datos debe ir siempre acompañado de una serie de archivos en **formato abierto y procesable**, que faciliten el acceso a los datos.

NOMBRAR ADECUADAMENTE COLUMNAS
Utilizar solo **caracteres ASCII en minúscula**. Los campos y sus especificaciones deben estar recogidas en el diccionario de datos que documenta el dataset. Tampoco deben usarse caracteres especiales, tildes o signos de puntuación. Los espacios deben ser sustituidos por guiones.

UTILIZAR UNA CODIFICACIÓN DE CARACTERES ESTANDARIZADA
Es recomendable emplear una **codificación de caracteres** internacionalmente reconocida, estandarizada o utilizada, como por ejemplo la codificación **UTF-8**.

ESTANDARIZAR VALORES DE DATOS
Para normalizar la estructura y los valores de los campos, es recomendable utilizar **vocabularios de referencia**. La estructura debe ser documentada en el **diccionario de datos**.

PUBLICAR DATOS COMPLETOS Y EVITAR VALORES AUSENTES
Ante la ausencia de datos en el conjunto, es necesario que el publicador especifique en el **diccionario de datos** la razón por la cual no están presentes. Para evitar confusiones, el publicador debe **marcar claramente los valores ausentes como valores nulos (NA)**.

EVITAR LA DUPLICIDAD DE REGISTROS
Estandarizar la recogida de datos y su almacenamiento, centralizando el proceso en un único sistema de información, de tal forma que las duplicidades sean fácilmente detectables y puedan ser eliminadas automáticamente.

ATRIBUTOS DE LA CALIDAD DE LOS DATOS

Exactitud	Completitud	Consistencia	Credibilidad	Actualidad	Accesibilidad
Conformidad	Confidencialidad	Eficiencia	Precisión	Trazabilidad	Comprensibilidad

Características de calidad de datos ISO/IEC 39013

PROPORCIONAR UNA CANTIDAD ADECUADA DE DATOS PARA FACILITAR SU ANÁLISIS
Los publicadores deben asegurar que se publica una **cantidad razonable de datos** para que haya suficiente contexto y los usuarios puedan obtener valor de su explotación.

FORMATEO DE VARIABLES DE FECHA Y HORA
Las fechas deben codificarse siempre utilizando el estándar internacional de referencia **ISO 8601**.

FORMATEO DE DATOS NUMÉRICOS
Utilizar como separador decimal el punto (internacionalización). Evitar separadores de millar. Valores negativos con signo (-). En columnas con valores enteros, no utilizar separadores decimales ni mezclar texto con valores numéricos.

EVITAR LA MEZCLA DE ESCALAS NUMÉRICAS
Intentar que la escala no varíe a lo largo del tiempo. En caso de ser necesario, proporcionar los datos en ambas escalas y documentar el cambio de escala.

EVITAR LA MEZCLA DE RANGOS EN UN MISMO CONJUNTO DE DATOS
Publicar los datos con el **mayor nivel de desagregación**. Si no es posible, mantener la consistencia en todos los valores de la variable.

INCORPORAR VARIABLES CON INFORMACIÓN GEOGRÁFICA
Publicar los datos con **coordenadas geográficas en dos columnas independientes: "latitud" y "longitud"**. Utilizar formatos específicos (SHP, KML) junto a otros que faciliten su reutilización (CSV, XLS).

EVITAR LA INCORPORACIÓN DE SUBTOTALES, TOTALES O AGRUPAMIENTOS
Presentar el **mayor nivel de desagregación posible** de los datos que contiene.

EVITAR LA FRAGMENTACIÓN DE DATOS Y DE DIFÍCIL LOCALIZACIÓN
Mejorar la **organización y etiquetado** de los contenidos, siendo necesario establecer conexiones entre los distintos conjuntos de datos.

ORGANIZAR ADECUADAMENTE LOS DATASETS DISPONIBLES
Organizar la publicación de **distribuciones atendiendo a formatos y dimensiones** (tiempo, geografía o temática).

Esta infografía pertenece a una serie de recursos divulgativos sobre la Guía de Calidad de los Datos Abiertos. En la siguiente, sigue aprendiendo sobre aspectos específicos de calidad aplicables en determinados formatos.

5. GUÍAS EXTERNAS - NORMAS UNE

*Las normas UNE de gestión y calidad del dato son un conjunto de estándares establecidos por la Asociación Española de Normalización (UNE) que se **enfocan en proporcionar directrices y mejores prácticas para garantizar la gestión eficaz y la calidad** de los datos en las organizaciones.*

*Estas normas buscan asegurar que los datos sean **precisos, completos, consistentes, accesibles y oportunos**, lo que es fundamental para la toma de decisiones informadas y el funcionamiento eficiente de aplicaciones basadas en datos.*

- **Privacidad de los datos**
- **Gestión del ciclo de vida de los datos**
- **Formación y concienciación**
- **Calidad de los datos**
- **Gestión de los metadatos**
- **Seguridad de la información**

5. GUÍAS EXTERNAS - NORMAS UNE

Buenas prácticas concretas en la creación y estructuración de los datos para el portal de Datos Abiertos son las siguientes:

Estandarización de Formatos: *Al combinar datos de múltiples fuentes, es importante estandarizar los formatos de datos para asegurar la coherencia y facilidad de uso.*

Limpieza y Normalización de Datos: *Realizar procesos de limpieza y normalización en los datos para corregir errores, eliminar duplicados y estandarizar la estructura de los datos.*

Privacidad y Seguridad

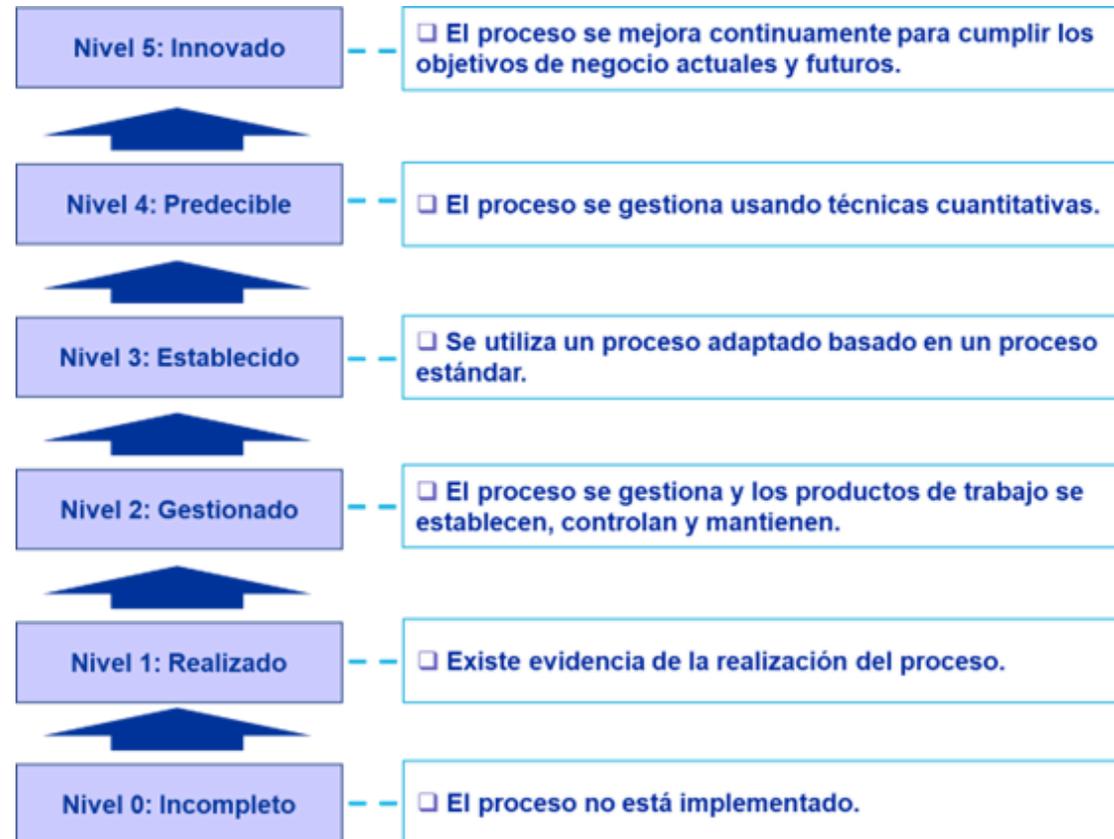
Validación y Verificación: *Realizar pruebas de validación y verificación para asegurar la calidad y precisión de los datos.*

Documentación y Metadatos: *Proporcionar metadatos claros que describan la fuente, el formato y el significado de los datos incluidos.*



5. GUÍAS EXTERNAS - NORMAS UNE

El modelo de evaluación de procesos establece que cada uno de los procesos tienen distintos niveles de capacidad o grados de mejora, siendo el nivel 0 el más básico y el nivel 5 el más avanzado:



5. GUÍAS EXTERNAS - NORMAS UNE

A continuación, se puede visualizar un *infográfico* con un resumen de las normas UNE para *asegurar la gestión y calidad del dato y las mejores prácticas*:

datos.gob.es



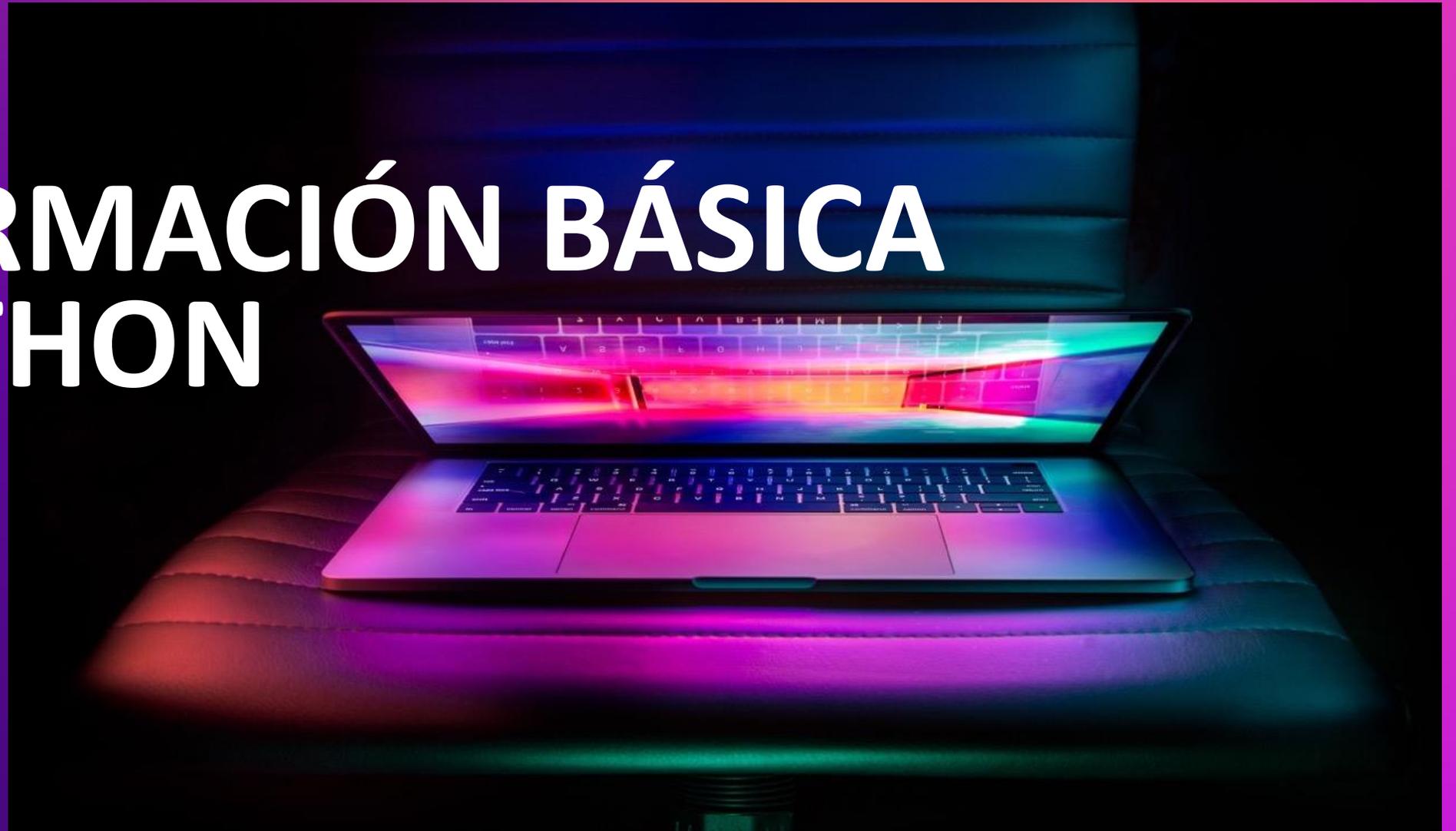
PREGUNTAS



ENCUESTA FINAL



FORMACIÓN BÁSICA PYTHON



6. ¿Qué es Python?

*Python es un lenguaje de programación de propósito general, es decir, se puede usar para casi cualquier aplicación. Es conocido por su **sintaxis clara y legible, que permite programar más rápido** en comparación con lenguajes como Java, C o R.*

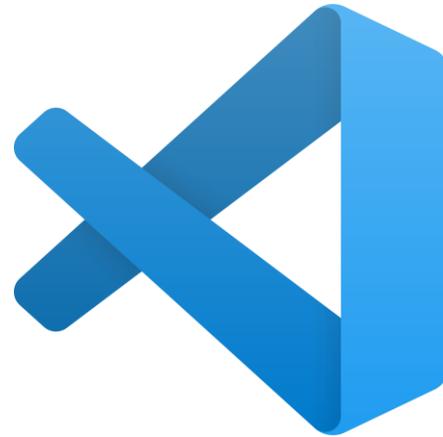
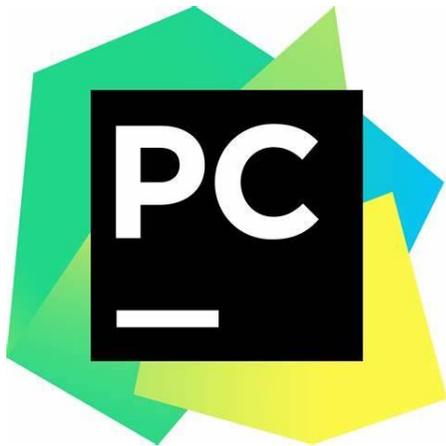
*Es un lenguaje de **código abierto**, por lo que, además de la biblioteca estándar, tiene una gran cantidad de bibliotecas y marcos de trabajo desarrollados por la comunidad.*

No es necesario declarar variables, como en otros lenguajes. Es un lenguaje orientado a objetos, e indentado (sangrado hacia la derecha), lo que hace más fácil su lectura.

```
1 from error_handler import ErrorHandler
2 import functions
3 import numpy as np
4
5 class RequestExtractor:
6     def __init__(self, c):
7         """
8         Constructor.
9         :param c: Configuration
10        """
11        self.config = c
12    def extract_params(self, d, app):
13        extracted_data = dict()
14        aux_list = list()
15        aux_used_columns = [x.lower() for x in self.config.used_columns]
16        for p in self.config.recommendation_req_params:
17            name = p[0]
18            if name.lower() in aux_used_columns:
19                mandatory = p[1]
20                try:
21                    if name.lower() == 'publishertime':
22                        extracted_data[name] = functions.get_time(d[name])
23                    elif name.lower() == 'browser':
24                        extracted_data[name] = functions.get_browser(d[name])
25                    elif name.lower() == 'os':
26                        extracted_data[name] = functions.get_os(d[name])
27                    else:
28                        extracted_data[name] = d[name]
29                except KeyError:
```

6. ¿Cómo ejecutar Python?

Podemos hacerlo desde la línea de comandos, o desde un intérprete. Existen muchos intérpretes: PyCharm, Spyder, Visual Studio Code, Jupyter... entre otros.



6. Conceptos importantes

Conceptos importantes que se desarrollarán durante la presentación:

VARIABLES Y OPERADORES:

- *Tipos de datos*
- *Operadores*

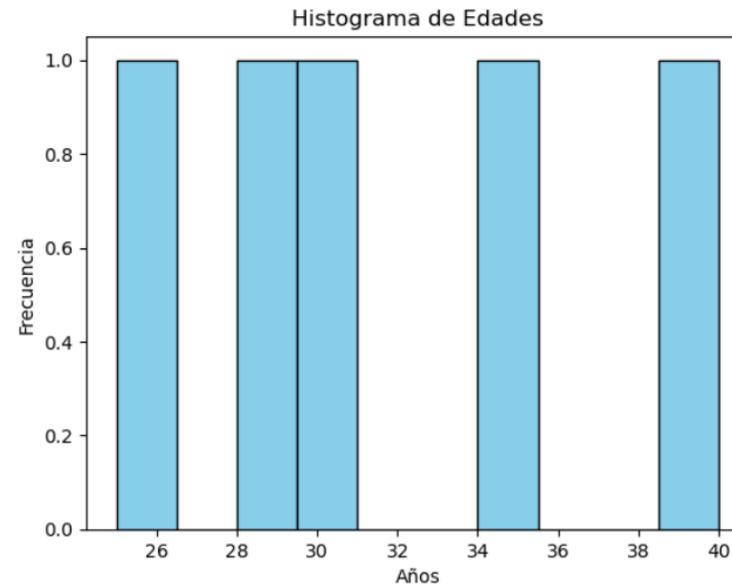
CONTROL DE FLUJO:

- *IF/Else*
- *Bucle For*
- *Bucle While*

DATAFRAMES:

- *Introducción a los Dataframes*
- *Como crear un Dataframe*
- *Exploración de un Dataframes*
- *Manipulación de Dataframes*
- *Limpieza y preparación de datos*
- *Visualización de datos*
- *Exportación de datos*

```
import matplotlib.pyplot as plt
plt.hist(df_completo['Años'], bins=10, color='skyblue', edgecolor='black')
plt.xlabel('Años')
plt.ylabel('Frecuencia')
plt.title('Histograma de Edades')
plt.show()
```



Gracias

