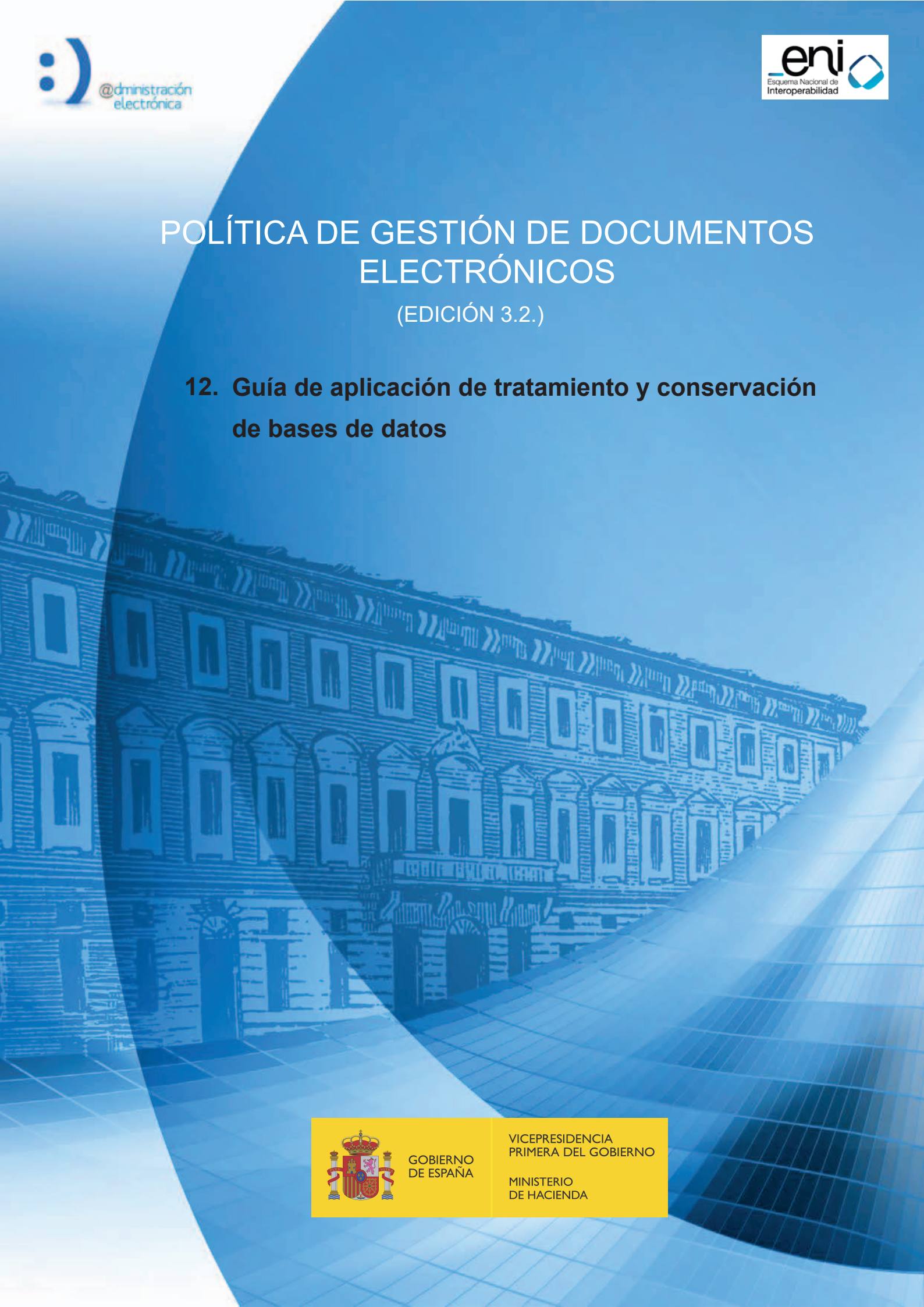
@dministración electrónica

eni Esquema Nacional de Interoperabilidad

POLÍTICA DE GESTIÓN DE DOCUMENTOS ELECTRÓNICOS

(EDICIÓN 3.2.)

12. Guía de aplicación de tratamiento y conservación de bases de datos



GOBIERNO DE ESPAÑA
VICEPRESIDENCIA PRIMERA DEL GOBIERNO
MINISTERIO DE HACIENDA

TÍTULO: 12. GUÍA DE APLICACIÓN DE TRATAMIENTO Y CONSERVACIÓN DE BASES DE DATOS

Elaboración y coordinación de contenidos:
Grupo de Trabajo de PGD-e, en el marco del Plan de Acción de Transformación Digital

Impulsa: Grupo de trabajo para la Coordinación de Archivos de MINHAC y Comisión Ministerial de Administración Digital (CMAD)

Dirige: Subdirección General de Documentación y Publicaciones de MINHAC
Responsable edición digital: Subdirección General de Documentación y Publicaciones
Edición electrónica (versión 1): noviembre de 2024
Disponible esta publicación en el Portal de Administración Electrónica (PAe):
http://administracionelectronica.gob.es/

Edita:
© Ministerio de Hacienda
Secretaría General Técnica
Subdirección General de Documentación y Publicaciones
NIPO: 225-24-132-9

El presente documento está bajo la licencia Creative Commons Reconocimiento-Compartir Igual versión 4.0 España.
Usted es libre de:
- Copiar, distribuir y comunicar públicamente la obra
- Hacer obras derivadas
Bajo las condiciones siguientes:
- Reconocimiento. Debe reconocer los créditos de la obra de la manera especificada por el autor o el licenciador (pero no de una manera que sugiera que tiene su apoyo o apoyan el uso que hace de su obra).
- Compartir bajo la misma licencia. Si altera o transforma esta obra, o genera una obra derivada, sólo puede distribuir la obra generada bajo una licencia idéntica a ésta.
Al reutilizar o distribuir la obra, tiene que dejar bien claro los términos de la licencia de esta obra.
Alguna de estas condiciones puede no aplicarse si se obtiene el permiso del titular de los derechos de autor
Nada en esta licencia menoscaba o restringe los derechos morales del autor.
Esto es un resumen legible por humanos del texto legal (la licencia completa) disponible en:

https://creativecommons.org/licenses/by-sa/4.0/deed.es

GUÍA DE APLICACIÓN:

TRATAMIENTO Y CONSERVACIÓN DE BASES DE DATOS

Texto-base y líneas de trabajo[1]

1. ANTECEDENTES 5

2. INTRODUCCIÓN 6

3. OBJETIVO Y ALCANCE 7

3.1. Qué NO incluye la Guía de Aplicación 8

4. ÁMBITO DE APLICACIÓN Y DESTINATARIOS 10

5. BASES DE DATOS: TIPOLOGÍAS Y ELECCIÓN. 11

5.1. Bases de datos relacionales 11

5.2. Bases de datos NoSQL/No Relacionales 13

5.2.1. Bases de datos clave/valor 13

5.2.2. Bases de datos documentales: 13

5.2.3. Bases de datos en orientadas a grafos. 14

5.3. Elección de la base de datos a utilizar según las necesidades. 16

6. RETOS PARA LA CONSERVACIÓN DE BASES DE DATOS 18

6.1. Dispersión de orígenes de datos 18

6.2. Incompatibilidades entre los diferentes productos de bases de datos 18

6.3. Complejidad 18

6.4. Dificultades para el procesamiento o manejo de ficheros de exportación correspondientes a bases de datos de gran tamaño 19

6.5. Volatilidad 19

6.6. Adaptación a la legalidad vigente 19

6.7. Seguridad 20

6.7.1. Gestión de riesgos. 20

6.7.2. Medidas de seguridad 22

6.7.3. Identificación de niveles de confidencialidad de documentos específicos 22

6.7.4. Gestión de las copias de seguridad 22

6.8. Protección de datos de carácter personal 23

[1] El texto que se inserta a continuación parte de la traducción de un artículo en inglés “Preserving Databases” contenido en la Nota técnica, de julio de 2021, elaborada para la Coalición de Conservación Digital. https://www.dpconline.org/docs/technology-watch-reports/2470-preserving-databases/file Sobre esta traducción se han producido aportaciones en las reuniones de este equipo de trabajo que complementan, modifican y suprimen parte del texto original para adaptarlo a los objetivos marcados.

7. FORMATOS DE ARCHIVO DE CONSERVACIÓN 24

7.1. Recomendaciones sobre formatos para facilitar la explotación y reutilización 26

7.2. Caracterización de formatos 27

8. HERRAMIENTAS DISPONIBLES PARA CONSERVACIÓN Y TRATAMIENTO DE BB.DD. 28

8.1. Exportación de bases de datos relacionales 28

8.2. Herramientas de procesamiento del lenguaje natural para caracterizar las BB.DD. 28

8.3. Procedimientos de conservación 29

8.3.1. Emulación 29

8.3.2. Migración 29

8.3.3. Copia 30

9. RECOMENDACIONES PARA LOS ORGANISMOS PRODUCTORES/GESTORES.. 31

9.1. Calendario de conservación y obligaciones legales 31

9.2. Garantía de la Calidad de los datos 32

9.3. Documentación y metadatos 33

9.3.1. Gestión de los metadatos 34

9.3.2. Metadatos de conservación dentro del modelo OAIS 35

9.3.3. Documentación de las bases de datos para su conservación. 38

9.3.3.1. Documentación del contexto 38

9.3.3.2. Consideraciones éticas. 39

10. DEFINICIONES Y ACRÓNIMOS 41

10.1. Definiciones 41

10.2. Acrónimos 49

11. REFERENCIAS 52

11.1. Estándares y buenas practicas 53

11.2. Bibliografía 57

1. ANTECEDENTES

1. Previamente a la redacción de este recurso por un *equipo de trabajo* dentro del *Grupo de Trabajo de la Política de Gestión de Documentos electrónicos*, se planteó el alcance y profundidad, conviniéndose que la conservación de las bases de datos entraba dentro de lo que, genéricamente, se entiende como "gestión del dato" y que, además de su conservación incluye otros aspectos como la reutilización. Tras varias reuniones y esquemas, compartidos luego en las reuniones del *Grupo*, se convino que, sin un marco regulatorio más amplio, el documento resultante sería demasiado genérico.
2. Al mismo tiempo, se consideró la necesidad, urgente, de abordar la conservación de las bases de datos, asociadas a procedimientos administrativos ya sea mediante gestión humana o por gestión administrativa automatizada, que contienen información no siempre volcada en documentos electrónicos o expedientes electrónicos que sean conformes el Esquema Nacional de Interoperabilidad (ENI).
3. Por el momento, no existe una solución técnica, asociada a la gestión de documentos electrónicos y el ciclo de vida de los documentos, capaz de gestionar la trasferencia de custodia de estas bases de datos como si fueran un "recurso archivístico", manteniendo la cadena de custodia al tiempo que asegura la disponibilidad de los datos.
4. Por ello, como primer paso, en lugar de abordar una "guía de aplicación para la gestión del dato" se planteó un objetivo mucho más práctico, centrado en la conservación de las bases de datos que se conservan en las distintas unidades del departamento, que se deben considerar como parte del patrimonio documental y, por tanto, tratar de manera acorde.

2. INTRODUCCIÓN

5. En la actualidad, una buena parte de la información generada por las administraciones públicas se encuentra estructurada en bases de datos. Su conservación es uno de los retos más importantes que se nos presenta a la hora de abordar una política de gestión de documentos electrónicos.
6. Este documento constituye una guía de aplicación sobre Tratamiento y Conservación de Bases de datos[2] en unidades que gestionen dichas bases de datos, especialmente aquellas consideradas con "valor patrimonial".
7. Según el Anexo II de la Política de Gestión de Documentos Electrónicos (PGDe) se consideran como bases de datos con *valor patrimonial* las que tienen un valor estadístico o las que recogen la información contenida en expedientes que se producen de una forma seriada, o ambas cosas. También se incluyen las bases de datos que pueden servir como fuente para emitir documentos solicitados por los ciudadanos, dando cumplimiento a los artículos 12 y 13 de la Ley 19/2013, de 9 de diciembre, de transparencia, acceso a la información pública y buen gobierno.
8. Los criterios que pueden ayudarnos para decidir si una base de datos tiene o no valor patrimonial son los siguientes:
    i. Permiten conocer los procesos de elaboración de leyes y reglamentos que afectan al organismo.
    ii. Permiten valorar la eficacia de las actividades del organismo.
    iii. Supervisan el funcionamiento interno del organismo.
    iv. Conservan datos significativos que permiten conocer la evolución política, económica y social de los territorios del Estado.
    v. Contienen datos necesarios para la protección de los derechos civiles, financieros, jurídicos, profesionales u otros derechos de los individuos, instituciones y de la misma institución.
    vi. Contemplan de manera significativa la información contenida en otros fondos o series documentales.
    vii. Responden a las necesidades del análisis estadístico y de la historia cuantitativa.

[2] Dado que no existe una Norma Técnica de Interoperabilidad (NTI) sobre la que realizar la guía, no se plantea la vinculación con una NTI ad hoc, aunque sí se haga referencias a otras NTI en vigor. Este particular se recogerá en el Alcance.

3. OBJETIVO Y ALCANCE

9. En esta guía de *tratamiento y conservación de bases de datos* se abordarán las condiciones y requisitos para la conservación de las bases de datos con el objetivo de garantizar su protección, conservación física y lógica de los datos y su contexto; así como asegurar su autenticidad, integridad, confidencialidad, disponibilidad y trazabilidad.

10. La conservación y tratamiento de las bases de datos se plantea como una obligación que deben asumir todos los organismos que las usan, recayendo esta obligación en las unidades tramitadoras, que han generado dichas bases o que han heredado las funciones de dichas unidades, tras cambios en la estructura administrativa.

11. Esta obligación parte de la consideración las bases de datos - asociadas a la gestión administrativa, y a la función de constancia y garantía de derechos de la Administración y los ciudadanos - como *docum*entos. Por tanto, dichas bases de datos deben considerarse como *Patrimonio Documental,* en los términos recogidos en la Ley 16/1985 de Patrimonio Histórico.

12. Esta consideración conlleva la necesidad legal de conservarlo para su uso por la administración, los ciudadanos o como fuente de investigación. La destrucción no autorizada de Patrimonio Documental entra dentro de la figura del *expolio* recogido en el artículo 4 de la Ley 16/1985 de Patrimonio Histórico, sancionado de manera acorde con los artículos 323 y 324 del Código Penal[3].

13. También hay que ser conscientes que la negligencia en la conservación de estos también está penada. El Código Penal sanciona en su artículo 413[4] la destrucción intencionada de documentos por autoridad o funcionarios público al que le esté encomendada su custodia. De manera similar, según el art. 264[5], es delito el borrado sin autorización de datos informáticos.

14. Según lo indicado en el apartado 5 de este documento, la conservación de una base de datos debiera abordarse desde el mismo momento de su análisis y diseño, el cual será determinante en la elección del tipo de base de datos a implantar y sus herramientas complementarias.

15. Durante la vida útil de la base de datos, además de su uso y administración, también habrá que plantear su conservación futura, asumiendo si los datos que gestiona o maneja tienen valores primarios (administrativos, jurídicos, fiscales, patrimoniales, etc.) y la relación de la base de datos con procesos administrativos que impliquen la generación de documentos en relación con los administrados o la propia actividad de la Administración a la que sirve.

16. También hay que plantearse, en la *valoración* de la base de datos, su potencial *utilidad de sustitución* – tanto por las facilidades diferenciales de consulta y análisis de la información que ofrecen las bases de datos en relación con documentos y expedientes, como por la posibilidad de usar dichas bases como respaldo de información contenida - de series documentales expurgadas, si se comprueba que contiene información similar.

---

[3] Libro II, Título XVI, Capítulo II: “De los delitos sobre el patrimonio histórico”.

4 Libro II, Título XIX; Capítulo IV: “De la infidelidad en la custodia de documentos y de la violación de secretos”.

5 Libro II, Título XIII, Capítulo IX: “De los daños”.

17. Finalmente, hay que considerar la posible *reutilización* de todos o parte de los *conjuntos de datos* contenidos en las bases de datos, por la propia administración o por terceros (ya sea mediante aprobación de acceso o datos abiertos), y los tratamientos informáticos que deben abordarse. En este proceso hay que garantizar la protección de datos personales y guardar la debida cautela para que la difusión de estos datos no suponga un perjuicio para la seguridad nacional.

18. Dentro de los procesos conservación hay que incluir la garantía de la continuidad en la actividad a la que sirven. Por ello, como paso integral de la gestión y conservación se deben tener en cuenta las necesarias medidas de seguridad recogidas en el Anexo II del Esquema Nacional de Seguridad (ENS) y considerar la adecuación de la base de datos, y los sistemas relacionados, al ENS, especialmente en lo que atañe a las *medidas de protección* que garanticen la accesibilidad e integridad de la base de datos, así como su confidencialidad, si fuera el caso.

19. Una vez que se tienen en cuenta las pautas y recomendaciones sobre implantación y uso, consignadas en este documento, es necesario considerar las estrategias de conservación - en la propia unidad productora o en la unidad destinaria tras una transferencia de custodia - y los formatos de conservación si se decide migrar los datos a alguno de los recogidos en este documento.

20. Es importante tener en cuenta que la base de datos deberá acompañarse de la documentación asociada en donde se documente el contexto de producción (propósito, metodología de carga de datos, tecnología asociada, etc.)

21. No siempre es necesario conservar la totalidad de una base de datos, sino que la conservación de los datos podría reducirse a un conjunto de tablas clave. Además de los datos, puede ser conveniente conservar el resultado de cruces, operaciones de *JOIN* (combinación) entre tablas, u otros procesamientos, que produzcan los datos relevantes a preservar, en vez de guardar toda la base de datos *en bruto*. En definitiva, de lo que se trata no es de preservar necesariamente toda la base de datos, sino el *contenido esencial* de la misma, en el cual se basan los actos administrativos, o bien el que se haya identificado en un proceso de calificación de la información de la base de datos.

22. Esta opción debiera incluir un proceso de certificación para aseverar que los datos recalculados o procesados que se han conservado provienen de la fuente original.

3.1. Qué NO incluye la Guía de Aplicación

23. Esta guía de aplicación del *de tratamiento y conservación de bases de datos* a la que da soporte esta guía, no debe considerarse como:

    viii. Una implantación específica. Las implantaciones específicas por llevar a cabo por cada organismo o unidad se diseñarán valorando la necesidad de aplicar requisitos adicionales en función de sus circunstancias específicas. En cualquier caso, estos requisitos serían diseñados e implantados bajo el principio de proporcionalidad y en concordancia con la normativa específica aplicable.

    ix. Una NTI (Norma Técnica de Interoperabilidad) de procedimiento de conservación de bases de datos. Las pautas para la conservación de bases de datos se definirán y normalizarán en la NTI de tratamiento y conservación de bases de datos. La realización de esta NTI está prevista en la Disposición Adicional Primera del ENI, punto 1.p.

x. Una regulación sobre los procedimientos de gestión del dato.

xi. Una regulación sobre reutilización de datos.

xii. Una regulación sobre seguridad de las bases de datos, incluido su posible cifrado. Éstas se aplican según lo establecido a tal efecto en el Esquema Nacional de Seguri*dad* (*ENS*) y el informe de buenas prácticas del CCN-CERT BP/24 *Recomendaciones de seguridad en Bases de datos*.

4. ÁMBITO DE APLICACIÓN Y DESTINATARIOS

24. El ámbito de aplicación de la guía de aplicación son los distintos organismos y unidades del Ministerio de Hacienda.

25. Dentro del ámbito de aplicación definido anteriormente, los destinatarios del contenido de la *Guía de aplicación de Tratamiento y Conservación de bases de datos* son los siguientes:

    i. Responsables de la ejecución de los diferentes procedimientos administrativos y procesos de gestión.

    ii. Desarrolladores de aplicaciones de gestión, conservación y archivo, de bases de datos, y documentos electrónicos.

    iii. Responsables de gestión, conservación y archivo de documentos electrónicos.

5. BASES DE DATOS: TIPOLOGÍAS Y ELECCIÓN.

26. Una base de datos es un conjunto formado por datos y sus relaciones, estructurados de manera que se permite el acceso a estos por distintas aplicaciones (BCS, 2013).

27. Un Sistema de gestión de bases de datos (*database management system*; DBMS) es la herramienta (o herramientas) de software que gestiona los datos en una base de datos (BCS, 2013).

28. Las bases de datos tienen diferentes formatos, estructuras, tamaños y niveles de complejidad. En los entornos *de producción*, las bases de datos pueden formar parte de sistemas de software que usan las bases de datos para proporcionar funcionalidad y acceso a los datos que gestionan.

29. Las estructuras de base de datos pueden variar desde tablas simples hasta conjuntos más complejos de esquemas, consultas, vistas, tablas y otra serie de elementos que, funcionando de manera conjunta, permiten que los datos puedan ser incorporados, borrados, cambiados, almacenados e interpretados por los usuarios.

5.1. Bases de datos relacionales

30. Dentro de los distintos tipos de bases de datos, las denominadas bases de datos "relacionales" son las más conocidas y usadas. (Freitas et al, 2009). Este modelo fue propuesto por Edgar Frank Codd en 1970, y su fundamento se basa en la *Teoría de Conjuntos* de la matemática.

31. Para su consulta se usa el *lenguaje estructurado de consulta* SQL (*Structured Query Language*), el cual es un *lenguaje declarativo* que está basado en el cálculo relacional de tuplas (o equivalentemente el álgebra relacional). SQL es un "lenguaje orientado a la transformación", y de cuarta generación (4GL), es decir, un lenguaje que usa relaciones para transformar unas entradas en las salidas deseadas.

32. En los lenguajes procedimentales de tercera generación se debían especificar todos los pasos que hay que dar para conseguir el resultado. Sin embargo, en SQL tan sólo deberemos indicar al Sistema de Gestión de Base de Datos (SGBD) qué es lo que queremos obtener - pero no cómo hacerlo-, y usando una sintaxis relativamente simple. La estructura de los comandos se basa en palabras del inglés estándar, como: *CREATE TABLE, SELECT, INSERT, UPDATE o DELETE*.

33. Las *bases de datos relacionales* sustituyeron a las *bases de datos de primera generación* (entre las que destacaban el *modelo jerárquico* y el *modelo en red*), las cuales se ejecutaban principalmente en "mainframes" y eran "navegacionales", puesto que requerían el uso de punteros para navegar de un *objeto* a otro. Estas bases de datos anteriores a las relacionales presentaban una serie de problemas, ya que: eran muy inflexibles en cuanto a las estructuras de datos y capacidades de búsqueda, carecían de un fundamento matemático, las transacciones operaban sobre registros individuales en vez de conjuntos, y las operaciones de consultas complejas requerían codificación avanzada.

34. Las bases de datos relacionales vinieron a resolver los siguientes problemas de las bases de datos de primera generación:

i. Incremento sustancial de la facilidad de uso de las bases de datos, sin requerir conocimientos avanzados de programación, y sin exigir al programador que sepa cómo navegar a través de una base de datos para obtener los valores.

ii. Las bases de datos anteriores no diferenciaban la vista de los datos (para programadores o usuarios) de su implementación física. En el modelo relacional se crea una vista lógica, siendo ésta un esquema conceptual que representa los datos de forma independiente de los detalles técnicos, y que no se ve afectado por operaciones como reorganizaciones de ficheros físicos que puedan realizarse en el nivel de implementación. Asimismo, si se elimina por ejemplo una entidad, las vistas lógicas que no se refieran a ella no deberán verse afectadas.

iii. Se facilita la tarea del administrador de Base de Datos, con comandos específicos, y mayores prestaciones de seguridad.

iv. Las bases de datos de primera generación carecían de un fundamento matemático, mientras que las bases de datos relacionales tienen una sólida base matemática.

v. Las bases de datos relacionales poseen controles de integridad de los datos, asegurando la precisión de los datos y la consistencia, y evitando las duplicaciones.

vi. Las consultas SQL de las bases de datos relacionales pueden en principio ser realizadas por usuarios finales en vez de programadores expertos, permiten acceder en un solo comando a conjuntos de datos en vez de registros individuales, y proporcionan un acceso no procedimental a los datos.

35. Sin embargo, las bases de datos relacionales también tienen algunas potenciales limitaciones o inconvenientes, como las siguientes:

    i. Al basarse en el concepto de tabla, presentan dificultad para representar relaciones complejas entre objetos.

    ii. Pueden mostrar problemas de escalabilidad, que afectan a la latencia en las consultas y el rendimiento, cuando se trabaja con volúmenes muy elevados de datos o tablas, o cuando se distribuyen en múltiples servidores, o cuando se quiere trabajar con millones de usuarios, por ejemplo, en las plataformas de comercio electrónico que requieren capacidades transaccionales a escala masiva.

    iii. Las bases de datos relacionales requieren un espacio de almacenamiento muy elevado, debido a su estructura en tablas con filas y columnas.

    iv. La flexibilidad del lenguaje de manipulación de datos de SQL, considerada generalmente como una ventaja, puede dar lugar sin embargo a la construcción de consultas SQL complejas e ineficientes, como joins multinivel, o consultas para recuperar todo el contenido de una tabla (“SELECT”), aplicadas sobre una tabla con un número muy elevado de registros, las cuales tendrían un tiempo de proceso muy elevado.

36. La recuperación de la información se hace, como hemos dicho, mediante consultas que operaran en el Lenguaje Estructurado de consultas (SQL). Las relaciones entre las tablas que constituyen el modelo de datos en que se estructura la abstracción de la realidad que es una base de datos relacional.

5.2. Bases de datos NoSQL/No Relacionales

37. Frente a las bases de datos relacionales, se están popularizando las denominadas **NoSQL** (**Not only** Structured Query Language) también denominadas como “no relacionales" para destacar el hecho de que pueden administrar altos volúmenes de datos no estructurados.

38. Las NoSQL se diferencian de las bases de datos relacionales en los siguientes aspectos:

    i. esquema prescindible,

    ii. desnormalización de los datos (para evitar la realización de joins o consultas combinadas),

    iii. gran capacidad de escalabilidad horizontal y replicación en múltiples servidores,

    iv. entre sus premisas no se encuentra garantizar las propiedades ACID6, si bien en ocasiones las bases de datos NoSQL permiten establecer por configuración el cumplimiento de las propiedades ACID.

39. Hay distintos tipos de bases de datos NoSQL entre las que destacan:

5.2.1. Bases de datos clave/valor.

40. Este tipo de bases de datos toman toman los datos como un conjunto de pares “clave-valor” en los que las claves cumplen funciones de identificadores únicos. Las *claves* o los valores pueden ser cualquier tipo de datos (y en general, objetos binarios). Los *objetos* o *registros* se localizan mediante la clave, y el objeto recuperado consiste en una colección opaca al motor de la base de datos, que puede albergar un conjunto de campos diferente y variable para cada registro (a diferencia de las bases de datos relacionales, cuyas tablas contienen una estructura de campos con tipos bien definidos).

41. Las *bases de datos clave/valor* brindan grandes capacidades de escalabilidad horizontal, flexibilidad para incorporar diferentes tipos de datos, y gran velocidad de respuesta a las consultas de lectura o escritura, aunque como contrapartidas, no permiten hacer *joins* o consultas complejas, y las búsquedas no están en general optimizadas, porque requieren examinar las colecciones completas proporcionadas con cada registro, y su funcionamiento sólo es óptimo cuando cada clave va asociada a un valor único (es decir, cuando cada registro está vinculado a un valor consistente en un campo simple, en vez de una colección o conjunto de campos). Como **ejemplos** de bases de datos clave/valor tenemos *Cassandra*, *Redis* y *Oracle NoSQL*.

5.2.2. Bases de datos documentales:

42. Estas bases de datos se utilizan para almacenar, recuperar y administrar datos semiestructurados. Almacenan cada registro/documento y sus datos asociados en una sola entrada de base de datos, a diferencia de las bases de datos relacionales, en las que la información de un objeto puede estar dispersa entre varias tablas. Estas bases de datos

[6] (Haerder et al, 2020), acrónimo formado por las iniciales de atomicidad (Atomicity), consistencia (Consistency), individualidad (Isolation) y durabilidad (Durability). Atomicidad (“todo o nada” en la ejecución del conjunto de pasos de las transacciones), Consistencia (Coherencia, visión única, e integridad de los datos), Aislamiento (ninguna operación puede afectar a otras) y Persistencia (una vez realizada la operación, los cambios serán permanentes, incluso aunque haya fallos del sistema o caídas del suministro eléctrico)

ofrecen mucha flexibilidad, escritura rápida, y consultas rápidas gracias a su gran capacidad de indexación.

43. Los documentos en las bases de datos documentales son autodescriptivos, ya que almacenan típicamente toda la información sobre un objeto, incluyendo tanto los valores de los datos como información sobre los tipos de datos. Las bases de datos documentales extienden el concepto de las bases de datos clave-valor, dado que las bases de datos documentales albergan conjuntos de clave-valor dentro de su unidad básica de trabajo, que son los documentos.

44. Los valores pueden pertenecer a una variedad de tipos y estructuras de datos, incluyendo: cadenas, números, *booleanos*, fechas, *arrays* y objetos. Los documentos se almacenan en formatos como JSON, BSON, YAML, y XML, siendo JSON el más común, por cuanto se trata del formato estándar de facto para serializar – almacenar - objetos. Con las bases de datos documentales, se evita la necesidad de tener una capa relacional intermedia, pudiendo almacenar los objetos directamente en JSON.

45. Además de la unidad básica de trabajo (los documentos), las bases de datos documentales presentan una funcionalidad que las diferencian de las bases de datos clave-valor, puesto que organizan los documentos en grupos denominados “colecciones”, pudiendo formularse una analogía entre éstas y las tablas de las bases de datos relacionales. Entre las BBDD documentales más reconocidas: MongoDb y CouchDB

5.2.3. Bases de datos en orientadas a grafos.

46. Las bases de datos orientadas a grafos son excelentes herramientas para trabajar con datos complejos, y especialmente en aquellos escenarios en los que el interés fundamental o primordial resida en las relaciones entre los datos. Esto es de gran utilidad cuando los datos que vamos a procesar tienen altos niveles de interrelación (ejemplos típicos son: las redes sociales, sistemas de detección del fraude, modelos médicos, sistemas de gestión de identidades y control de acceso, motores de recomendación en tiempo real, grafos de conocimiento en aplicaciones de inteligencia artificial, planificación de cadenas logísticas, evaluación del riesgo, o topologías de red).

47. Las bases de datos orientadas a grafos tienen un sólido fundamento teórico (basado en la Teoría de Grafos de la matemática), a diferencia de otras bases de datos NoSQL.

48. Las bases de datos orientadas a grafos tienen tres componentes principales. El primero son los *nodos*, que representan una entidad (por ejemplo, un producto o un cliente). El segundo componente son los *vértices o relaciones*, que conectan a los nodos del grafo, y pueden ser dirigidos o no dirigidos. El tercero son las *propiedades*: conjuntos de parejas *nombre: valor* que se pueden añadir tanto a los nodos como a las relaciones, para proporcionar información significativa de contexto.

49. Si bien las bases de datos relacionales son capaces de modelar estos grafos o redes, mediante el uso de claves foráneas o “*self-joins*” (consultas de combinación de una tabla consigo misma), presentan problemas de rendimiento al trabajar con grafos de gran tamaño, y no poseen un lenguaje de consultas con una sintaxis expresiva o intuitiva para trabajar con grafos, dándose además la circunstancia de que SQL no tiene la capacidad de recorrer o atravesar gráficos cuya profundidad sea desconocida o no acotada.

50. Las bases de datos orientadas a grafos tienen un rendimiento excelente para consultar datos con un número elevado de relaciones, dado que el tiempo necesario para atravesar

los nodos vecinos de uno determinado es independiente del tamaño del grafo (X – antes-Twitter - es un ejemplo de uso de estas bases de datos en el *backend*). Otra ventaja de las bases de datos orientadas a grafos es la agilidad o flexibilidad para hacer cambios en el modelo a medida que evolucionan los requisitos de negocio, sin tener que hacer una reestructuración significativa del modelo de datos que pudiera tener un impacto en la funcionalidad o en el tiempo de implementación.

51. La base de datos orientada a grafos más conocida es Neo4j. Lenguajes como SPARQL, Cypher o Gremlin están optimizados para las operaciones de recorrido de gráficos, con una sintaxis intuitiva que evita los *joins* recursivos de las bases de datos relacionales. Como inconvenientes, las bases de datos orientadas a grafos no son idóneas o eficientes para los escenarios transaccionales en los que se realizan consultas de grandes volúmenes de datos, ni para la minería de datos, ni para las bases de datos que contengan pocas relaciones, y, por otro lado, no disponen de un lenguaje de consulta común a los diferentes fabricantes, lo cual dificulta las migraciones y los procesos de aprendizaje de los programadores.

52. Otro tipo de bases de datos NoSQL serían:

    i. Bases de datos multivalor

    ii. Bases de datos orientadas a objetos

    iii. Bases de datos tabulares

    iv. Bases de datos de Arrays

53. En relación con los tipos de bases de datos NoSQL, merece la pena destacar que la línea que los separa es un tanto difusa, puesto que gran parte de estas bases de datos abarcan varias categorías. Así, por ejemplo, Cassandra se puede considerar como una base de datos híbrida: clave-valor y columnar o tabular.

54. Frente a la estructura de tablas de las bases de datos relacionales, en las no-relacionales tenemos estructuras no-tabulares, bases de datos de texto plano (*flat database*) o conjuntos de estructuras de datos múltiples y complejas. A la hora de procesar una *transacción*[7] de bases de datos, todas comparten cuatro propiedades clave que, como se ha dicho, se resumen con el acrónimo “ACID”,

55. Las bases de datos NoSQL generalmente hacen concesiones al flexibilizar algunas de las propiedades ACID de las bases de datos relacionales para un modelo de datos más flexible que puede escalar horizontalmente. Así, NoSQL debilita los requisitos de: consistencia inmediata, datos actualizados en todo momento, y exactitud de respuesta, con objeto de ganar escalabilidad y una recuperación más rápida de los datos.

56. Para la recuperación de datos de una base de datos NoSQL se requieren nuevos conceptos, y puede ser necesario un nuevo lenguaje de consultas. Como ejemplos de estos lenguajes tenemos Gremlin o SPARQL. Se debe tener en cuenta que algunos consideran el nombre “NoSQL” como poco afortunado, y no debe tomarse ese “no” como modificador negativo de SQL, sino como un acrónimo de “*not only*”, que tiene esa matización de “no solo”; dado que no es la ausencia de uso de SQL la que caracteriza a estas bases

[7] En este tipo de proceso el trabajo se divide en operaciones individuales, indivisibles denominadas “transacción”. Estas operaciones se oponen a los procesos masivos donde uno o más programas procesan conjuntos de datos (batch) sin intervención, apenas, de un operador externo.

de datos (de hecho, muchas de ellas ofrecen también la posibilidad de realizar consultas con lenguaje SQL), sino la divergencia respecto al modelo de datos relacional y transaccional.

5.3. Elección de la base de datos a utilizar según las necesidades.

57. A la hora de decidirse por el uso de bases NoSQL o Relacionales, hay que considerar algunos factores:

i. Cuando los datos deben ser consistentes sin dar posibilidad al error, conviene utilizar una base de datos relacional.

ii. Se optará por NoSQL cuando la adaptación a las demandas de prestaciones crecientes se implemente mediante la aproximación de la escalabilidad horizontal, usando un conjunto de nodos de menor rendimiento que permiten alcanzar las mismas prestaciones que se obtendrían, en el caso de una arquitectura de servidor único, sustituyendo la máquina por otra más potente.

iii. En los sitios web distribuidos a gran escala, que poseen servidores en diferentes lugares del mundo para proporcionar tiempos de respuesta reducidos a nivel global, las transacciones ACID se vuelven difícilmente sostenibles. Ello es debido fundamentalmente a que, de acuerdo con el teorema CAP (Consistency, Availablilty, Partition Tolerance) establecido por Eric Brewer en el año 2000, en los sistemas distribuidos solo se pueden obtener como máximo 2 de esas 3 propiedades a la vez. Si queremos que un sistema de base de datos distribuida siga estando disponible a los usuarios ante fallos entre dos segmentos de red del sistema distribuido (o pérdida de mensajes arbitrarios en redes imperfectas), debemos sacrificar la Consistencia (es decir, es posible que no todos los usuarios de un sistema distribuido tengan una vista idéntica de los datos en todo instante).

   En este sentido, las bases de datos NoSQL, sustituyen la consistencia estricta por el concepto de consistencia eventual, que implica que todos los usuarios (por ejemplo, en una aplicación de red social o de comercio electrónico) acabarán viendo los mismos datos al cabo de un tiempo, pero no de manera instantánea.

   En los *sitios web* de escala masiva, que disponen de nodos en todo el planeta, se da prioridad a la disponibilidad del servicio y a la latencia, o tiempo de respuesta reducido; siendo así que la obtención de una Consistencia Estricta resulta inasumible, puesto que implicaría la necesidad de que todo cambio en bases de datos se propague de manera síncrona e inmediata a múltiples nodos, incluyendo aquellos situados a miles de kilómetros y en otros continentes, y ello repercutiría negativamente en el tiempo de respuesta del servicio.

   Las bases de datos NoSQL, que no tienen como objetivo de diseño garantizar las propiedades ACID como las bases de datos Relacionales, y que implementan Consistencia Eventual en vez de Consistencia Estricta, resultan muy apropiadas en estos contextos de aplicaciones online masivas de alta disponibilidad global.

   Varias bases de datos NoSQL (como Cassandra, MongoDB y otras) poseen funcionalidades de Consistencia Ajustable por configuración, que permiten establecer el número de réplicas de datos que deben responder a una petición para que la operación de lectura o escritura sobre una base de datos se considere completa,

y de esta manera se pueden alcanzar distintos equilibrios entre: rendimiento de lectura, rendimiento de escritura, y consistencia.

iv. Para tiendas online con motores de inteligencia complejos, o aplicaciones web que operan a escala masiva con servidores instalados por todo el mundo, y en las que la rapidez de respuesta global y la disponibilidad sean más importantes que la consistencia inmediata, se recomienda NoSQL.

v. Cuando las estructuras de datos que manejamos son variables, heterogéneas, y no estructuradas o semiestructuradas, se recomienda NoSQL.

vi. Si tenemos que hacer análisis de grandes cantidades de datos en modo lectura con una baja latencia, se recomienda NoSQL

Por los cinco últimos puntos, las bases de datos NoSQL son, generalmente, las más usadas en las aplicaciones de Big Data.

58. Es muy importante en este punto insistir en que, aunque parece que en estos momentos sería conveniente plantearse la migración a bases de datos NoSQL, debemos tener muy en cuenta antes de tomar esta decisión si las características de nuestra base de datos necesitan una base de datos NoSQL o relacional. Como veremos más adelante, al hablar del proyecto *E-Ark* y del formato SIARD, la tecnología permite guardar bases de datos relacionales en formato no relacional y mediante importación volver a recuperar el funcionamiento inicial.

6. RETOS PARA LA CONSERVACIÓN DE BASES DE DATOS

59. La conservación de una base de datos se complica conforme esta crece en tamaño, complejidad y frecuencia de modificación.

6.1. Dispersión de orígenes de datos

60. Hay que tener en cuenta la existencia de múltiples fuentes de datos con diferentes grados de organización y agregación que deben abordarse de manera coherente:

    i. Consolidación y transformación de múltiples fuentes de datos, con diferentes grados de actualización (habiendo bases de datos departamentales y centralizadas, operacionales y analíticas, etc.).

    ii. Existencia de documentos almacenados fuera de la base de datos (con un enlace a su ubicación en un sistema de almacenamiento), siendo necesario garantizar la conservación de estos, no sólo de la base de datos.

    iii. Es muy habitual que una base de datos sea utilizada por varios sistemas de información, por lo que es necesario guardar información de configuración, que identifique los sistemas o aplicaciones vinculados a una determinada base de datos.

6.2. Incompatibilidades entre los diferentes productos de bases de datos

61. Cada suministrador de sistemas de gestión de bases de datos tiene tipos de datos y formatos de exportación que son específicos, y presentan variaciones en sus implementaciones de SQL respecto a los estándares, lo cual conlleva que en la práctica es infrecuente que la información exportada a partir del sistema de un suministrador pueda funcionar en otro diferente (o a veces incluso en versiones distintas de un mismo producto). De ahí la necesidad de emplear formatos de conservación completamente estándares e independientes de suministradores, evitando el SQL nativo, y realizando en su caso transformaciones o conversiones de datos previamente a la exportación.

6.3. Complejidad

62. La conservación de una base de datos implica muchas tareas simultaneas (RDB SIARD, 2019):

    i. Conservación de la información digital en un formato específico de datos.

    ii. Conservación de la estructura de la base de datos y la estructura lógica de la información.

    iii. Conservar objetos complejos o de gran tamaño asociados a la base de datos.

    iv. Empaquetar datos, estructura y documentación relacionada en “paquetes de archivo” que garanticen su conservación a largo plazo. La conservación de la documentación relacionada es importante para conocer el contexto en que opera la base de datos y la procedencia de los datos.

6.4. Dificultades para el procesamiento o manejo de ficheros de exportación correspondientes a bases de datos de gran tamaño

63. Si bien formatos estándar como XML o CSV no tienen un límite teórico de tamaño, en la práctica se encuentran limitaciones en cuanto a su usabilidad, porque por ejemplo la mayoría de ordenadores para usuarios finales presentan problemas para "parsear" o analizar ficheros XML de un tamaño superior a 200 MB, o las aplicaciones de hojas de cálculo no permiten abrir ficheros CSV que superen un número elevado de filas y columnas preestablecido según aplicación y versión.

6.5. Volatilidad

64. La información contenida en las bases de datos puede ser volátil, según se produzcan cambios o actualizaciones frecuentes (Thomson, 2016). Esta característica conlleva una serie de problemas y retos añadidos:

    i. La necesidad de realizar frecuentemente copias de respaldo, para evitar la pérdida de datos, implica un aumento de los requisitos de espacio de almacenamiento de datos.

    ii. El almacenado y la recuperación de la información se complican según el número de columnas y filas, la complejidad de las relaciones entre las tablas y la cantidad de datos que se gestionen.

    iii. La volatilidad de los datos también supone un reto para la valoración, especialmente en entornos cambiantes en donde, en lugar de añadir constantemente información, se producen constantes cambios o eliminación de datos pre-existentes. Esto aconsejaría revisar o actualizar el proceso de calificación en función de estos cambios o eliminaciones.

6.6. Adaptación a la legalidad vigente

65. La reutilización y la conservación de datos personales se verán afectados si una base de datos, o los equipos en los que reside, están afectados por diferentes marcos regulatorios o jurisdicciones. Por ejemplo, la AGE se puede ver afectada por normativa europea. También habría que ver cuándo se necesita tener en cuenta los distintos marcos regulatorios autonómicos o la interacción con la Administración Local.

66. Estos marcos regulatorios se deben reflejar en políticas de seguridad y privacidad, dentro de la organización, junto con calendarios de conservación y otras medidas organizativas que deben adoptar los archivos para conservar las bases de datos y asegurar que sean accesibles para consultas futuras.

67. Se cumplirá con la legislación vigente en materia de protección de datos de carácter personal que resulte de aplicación, en concreto el REGLAMENTO (UE) 2016/679 DEL PARLAMENTO EUROPEO Y DEL CONSEJO de 27 de abril del 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos y por el que se deroga la Directiva 95/47/CE (Reglamento General de Protección de Datos RGPD); Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y Garantía de los Derechos Digitales (LOPDGDD); así como las disposiciones de desarrollo de las normas anteriores o cualesquiera otras aplicables en materia de protección de datos que se encuentren en vigor.

68. De la misma manera y de obligado cumplimiento tanto para Administraciones Públicas y prestadores de servicios, y con el fin de mantener la seguridad de los tratamientos de datos deberá aplicarse el Real Decreto 311/2022, de 3 de mayo, por el que se regula el Esquema Nacional de Seguridad (ENS), con el objetivo de establecer las medidas que deberán implantarse para evitar su pérdida, alteración o acceso no autorizado adaptando los criterios de determinación del riesgo en el tratamiento de los datos.

69. Asimismo, le será de aplicación el Real Decreto-ley 24/2021, de 2 de noviembre por el que se modifica las previsiones recogidas en la Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público.

70. Será de aplicación también el Real Decreto 4/2010, de 8 de enero, por el que se regula el Esquema Nacional de Interoperabilidad (ENI) en el ámbito de la Administración Electrónica, que establece un conjunto de criterios y recomendaciones que regulan la manera en la que las Administraciones Públicas deben tratar la información que comparten con otros sistemas de la propia Administración.

71. Adicionalmente deberá tomarse en cuenta toda norma posterior que le sea de aplicación dentro del ámbito de la presente guía.

6.7. Seguridad

72. De igual manera al punto anterior, entre otros, hay que tener en cuenta que:

    i. De acuerdo con el artículo 17 del ENS, el acceso controlado a las BBDD deberá estar limitado a los usuarios, procesos, dispositivos u otros sistemas de información, debidamente autorizados, y exclusivamente a las funciones permitidas.

    ii. Así como también se recoge en el artículo 22 del ENS la Protección de información almacenada y en tránsito establece las condiciones necesarias para una adecuada protección de la información que se encuentra en dispositivos de almacenamiento, así como aquella que se encuentre en soporte no electrónico.

    iii. El CCN-CERT en su guía BP/24: Recomendaciones de seguridad en bases de datos, indica que las cuentas de los usuarios tienen que estar debidamente protegidas, con identificadores únicos, contraseñas robustas y deben bloquearse tras un determinado número de intentos fallidos de acceso.

6.7.1. Gestión de riesgos.

73. Por riesgos entenderemos tanto las amenazas externas o internas, contingencias, accidentes y cualquier situación que tenga como resultado de su aparición un impacto negativo. El riesgo es tanto la exposición a lo incierto como las desviaciones potenciales a lo que tenemos planeado.

74. La *gestión de riesgos* es tomar las decisiones necesarias para reducir los efectos de la materialización de un riesgo a un nivel tolerable. La gestión de riesgos pasa por las siguientes etapas:

    i. Identificar el contexto organizativo donde se deben gestionar los riesgos

    ii. Documentar el marco regulatorio

    iii. Identificar actividades, recursos y sus responsables

    iv. Identificar los riesgos

v. Individualizar y evaluar los riesgos sobre los recursos y actividades

vi. Diseñar y aplicar medidas para prevenir, afrontar y gestionar los riesgos

75. Según el art. 7 del ENS la gestión de seguridad se basa en el análisis y la gestión de riesgos. La gestión de los riesgos permitirá el mantenimiento de un entorno controlado, minimizando los riesgos a niveles aceptables. La reducción a estos niveles se realizará mediante una apropiada aplicación de medidas de seguridad, de manera equilibrada y proporcionada a la naturaleza de la información tratada, de los servicios a prestar y de los riesgos a los que estén expuestos.

76. En este contexto la gestión documental lleva implícito un ejercicio de gestión de riesgos donde el objetivo es transformar la incertidumbre en lo que respecta al mantenimiento de la usabilidad y seguridad de objetos digitales en riesgos medibles y gestionables.

77. Dentro de la gestión de riesgos tenemos la metodología MAGERIT y las siguientes herramientas

i. PILAR y uPilar (microPilar)

ii. El Marco de Auditoría de Datos (DAF)

iii. Nestor: Catalogue of Criteria for Trusted Digi¬tal Repositories

iv. Trusted Repository Archiving Checklist (TRAC) / ISO 16363[8]

v. DRAMBORA[9]. Digital Repository Audit Method Based on Risk Assessment

78. La gestión de riesgos incluye la identificación de los documentos esenciales del organismo y las políticas para asegurar su accesibilidad e integridad y recuperación de servicio ante incidentes de seguridad. Para la gestión de la seguridad de nuestras bases de datos tendremos que evaluar si las debemos considerar incluidas dentro del repertorio de “documentos esenciales”. Este aspecto está contemplado dentro del anexo X de la Política de Gestión de Documentos Electrónicos del Ministerio de Hacienda.

79. Dados los costes que supone la instalación de sistemas de respaldo y de copias de seguridad, sería necesario, por parte de los responsables de la información, identificar e inventariar toda aquella documentación que se produzca en los sistemas, destacando aquella que se considera “documentación esencial” e indicar las necesidades de disponibilidad tras un desastre.

80. Un almacenamiento digital de confianza requiere una serie de previsiones técnicas y organizativas, así como el alineamiento con los propósitos que llevaron al almacenamiento de la información y las especificaciones con las que se operó y mantenerlas durante tanto tiempo como sea preciso.

---

[8] Las métricas TRAC proporcionan las bases para lista de comprobación Trusted Digital Repository (TDR), aprobada como un estándar para la auditoría y certificación de repositorios Digitales de confianza. TRD se publicó en 2011 como estándar CCSDS (Recommended Practice 652.0-M-1 que se ha visto en un apartado anterior ya que este comité también impulsó su desarrollo. El estándar ISO 16363, que es una revisión de TDR (originado a partir de las métricas TRAC) se aprobó en febrero de 2012.

[9] DRAMBORA surgió de la toma de conciencia sobre los retos tecnológicos, organizacionales y metodológicos que supone la configuración de un repositorio. Si estos retos se consideran como riesgos evitables o asumibles se puede empezar a ver el modo de superarlos. Por ello el primer paso es identificar los riesgos que afronta (análisis de riesgo) y luego saber cómo gestionarlos (gestión de riesgos). Esta metodología para asegurar la conservación a largo tiempo, que desarrollaron las bibliotecas, es aplicable a los archivos.

DRAMBORA no es una herramienta para certificar la conformidad con OAIS. Es una herramienta de gestión y autoevaluación diseñada para evaluar la capacidad de una organización para conservar sus materiales digitales.

6.7.2. Medidas de seguridad

81. El ENS, en su anexo II, trata de las medidas de seguridad para el cumplimiento de los principios y requisitos mínimos de seguridad, siendo estas medidas proporcionales a las dimensiones de seguridad relevantes en el sistema a proteger y la categoría de seguridad del sistema de información.

82. Por ello se tratan a continuación dos de los principales aspectos que afectarían, especialmente, a la conservación de bases de datos, entendiendo que – previamente – se han adoptado otras medidas en los marcos organizativo y operacional:

    i. la identificación de los niveles de confidencialidad de la información, según la política de seguridad de la información adoptada

    ii. la gestión de las copias de seguridad, acorde a la categoría de seguridad del sistema de información, para garantizar la integridad y accesibilidad de la documentación en caso de desastre.

6.7.3. Identificación de niveles de confidencialidad de documentos específicos

83. Dentro de la gestión documental es fundamental identificar y localizar aquella documentación que pudiera tener algún tipo de clasificación de seguridad o que, por su sensibilidad, pudiera implicar un perjuicio para los intereses nacionales. El ENS denomina a este proceso “calificación”. Para calificar la información [mp.info.2] se estará a lo establecido legalmente por las leyes y tratados internacionales de los que España es miembro y su normativa de aplicación cuando se trate de materias clasificadas.

84. El **responsable de cada información** seguirá los criterios determinados en la Política de Seguridad de la Información para asignar a cada información el nivel de seguridad requerido, y será responsable de su documentación y aprobación formal.[mp.info.2.4]

6.7.4. Gestión de las copias de seguridad

85. De acuerdo con indicado en el apartado anterior, la categoría y riesgos de un sistema determinarán unas determinadas medidas de protección. Dentro de las medidas de protección de la información [mp.info] están las copias de seguridad [mp.info.6].

86. El art. 26 del ENS indica que los sistemas dispondrán de copias de seguridad y se establecerán los mecanismos necesarios para garantizar la continuidad de las operaciones en caso de pérdida de los medios habituales.

87. El ENS indica que se realizarán copias de seguridad que permitan recuperar datos perdidos, accidental o intencionadamente [mp.info.6.1]. La periodicidad y los plazos de retención de estas copias de seguridad se determinarán en la normativa interna de la organización relativa a copias de seguridad. Según la categoría de nuestro sistema y los riesgos así serán los procedimientos de respaldo [mp.info.6.2] establecidos, variando:

    i. La frecuencia de las copias.

    ii. Los requisitos de almacenamiento en el propio lugar.

    iii. Los requisitos de almacenamiento en otros lugares.

    iv. Los controles para el acceso autorizado a las copias de respaldo.

88. Si fuera el caso, las copias de seguridad de nuestras bases de datos deben conservarse, con las medidas adicionales de deslocalización y multiplicidad de copias (tanto en número

como en formato) ya que será el punto de partida tras un incidente, emergencia o desastre. En función de un análisis de riesgos, adaptado a las particularidades del organismo, podrán implementarse otras medidas de protección basadas en un programa de copias de seguridad, replicación y deslocalización.

6.8. Protección de datos de carácter personal

89. El ENS en su medida [mp.info.1] Datos personales, determina que “Cuando el sistema trate datos personales, el responsable de seguridad recogerá los requisitos de protección de datos que sean fijados por el responsable o por el encargado del tratamiento, contando con el asesoramiento del DPD (Delegado de Protección de Datos), y que sean necesarios implementar en los sistemas de acuerdo a la naturaleza, alcance, contexto y fines del mismo, así como de los riesgos para los derechos y libertades de acuerdo a lo establecido en los artículos 24 y 32 del RGPD, y de acuerdo a la evaluación de impacto en la protección de datos, si se ha llevado a cabo”.

90. Así pues, de acuerdo con lo que recoge el artículo 24 del RGPD corresponde al responsable del tratamiento la aplicación de medidas técnicas y organizativas que garanticen el correcto tratamiento de los datos. En la misma línea el artículo 32 del referido Reglamento se recoge una serie de medidas técnicas y organizativas las cuales deben ser aplicadas por el responsable del tratamiento entre las que se incluyen la seudonimización y cifrado de datos personales, capacidad para garantizar la confidencialidad, integridad, disponibilidad, capacidad para restaurar la disponibilidad y el acceso a los datos personales, entre otros.

91. El RGPD establece una serie de derechos que son de obligado cumplimiento respecto al tratamiento de los datos, los cuales dotan de un mayor de cumplimiento de la seguridad de estos, tales como el derecho de acceso, rectificación, supresión, oposición, limitación, así como el deber de informar sobre el tratamiento de los datos el cual se encuentra ligado a la publicación del Registro de Actividades del Tratamiento recogido en el artículo 30 del RGPD.

92. Debe tenerse en cuenta que, para garantizar la eficacia de la seguridad de los datos personales, en caso de que sucediese una brecha de seguridad de los mismos, esta debe comunicarse, siempre y cuando se considere necesario y sin dilación indebida, a la autoridad de control de conformidad con el artículo 33 del RGPD.

7. FORMATOS DE ARCHIVO DE CONSERVACIÓN

93. Dada la variedad de soluciones aplicadas, no hay un formato de fichero de base de datos que se pueda considerar que garantice la conservación futura o que sea el más adecuado para la conservación.

94. A la hora de elegir un formato de conservación la decisión variará según aquellos aspectos y funcionalidades que quieran conservarse para su uso futuro. Dentro de sus recomendaciones, la Biblioteca del Congreso de los Estados Unidos (2020-2021) solo indica que "debe conservarse el conjunto completo del contenido de la base de datos".

95. Para las bases de datos consistentes en recursos textuales, quizás la alternativa más sencilla sea el uso de formatos de archivo basados en texto (que guarden tanto los datos de la base de datos como estructuras de tablas, relaciones, *triggers*, procedimientos, y vistas de tablas) considerados como estándares "abiertos" que impliquen neutralidad tecnológica e independencia de los proveedores de aplicaciones, al tiempo que garantizan una accesibilidad óptima a los datos (Thomson, 2016).

96. Sin embargo, podemos tener bases de datos orientadas a la conservación y gestión de objetos digitales. En este caso se plantea el doble reto de conservar tanto la base de datos como los objetos asociados en soportes de conservación. Un formato interesante es el *formato de imagen FITS*, utilizado por la NASA, y que ha sido implementado como formato de conservación digital por la Biblioteca Vaticana y por Patrimonio Nacional en España. Este tipo de archivos, además de imágenes convencionales, puede almacenar datos de diversas clases como espectros electromagnéticos (imágenes de rayos x y de infrarrojos), mapas de bits, texto ASCII, matrices multidimensionales, tablas binarias, palabras clave que proporcionan información descriptiva sobre los datos que se almacenan, etc. Esto es posible gracias a que un mismo fichero, además de la Primary Header (Cabecera primaria), posee múltiples extensiones (o bloques de datos) con información diferente de un mismo objeto en cada una de ellas.

97. Junto con las tablas y su estructura será necesario incluir la documentación relacionada sobre el diseño y la evolución de las bases de datos. Véase apartado 6.2 de este documento.

98. En la tabla que sigue se describen los formatos de fichero que pueden utilizarse para garantizar la conservación y el acceso futuro.

Formato de fichero	Extensión	Observaciones
Valores separados por un carácter de delimitación	.csv .tsv .txt	Los ficheros de *valores separados por un delimitador* almacenan los datos en filas y columnas de datos usando caracteres de separación como comas, tabuladores o símbolos. Los Valores Separados por Comas (.CSV) y los Valores Separados por Tabulador (.TSV) son dos formatos de archivo usados habitualmente para exportar datos desde las bases de datos. Por ejemplo, una base de datos relacional puede exportarse como uno o varios ficheros .CSV (en este último caso, se suele comprimir el conjunto de ficheros en un solo archivo ".ZIP"). Aunque, mediante el uso de herramientas especializadas (querycsv.py o similares), se pueden realizar consultas o generar informes, la práctica más habitual es abrir estos ficheros en programas de hoja de cálculo o mediante herramientas de manipulación de datos como OpenRefine. No obstante, esta manipulación puede dar problemas según la potencia de proceso del equipo en que se manipulen. Los ficheros que contienen valores separados por un delimitador suelen conservar los datos contenidos en las tablas, pero no otros aspectos de una base de datos como fórmulas, interfaces de usuario, utilidades de informe y relaciones complejas. Antes de optar por uno de estos formatos habrá que considerar cuidadosamente qué aspectos de la base de datos deben conservarse o documentarse. Este tipo de soluciones .CSV, .TSV son únicamente recomendables para bases de datos pequeñas y con pocas tablas. Algunas herramientas de conservación de bases de datos como *Chronos* o *FullConvert* permiten la exportación de bases de datos a formatos .CSV
SIARD 1.0	.siard	Software Independent Archiving of Relational Databases (SIARD) es un formato de conservación, no vinculado a ningún proveedor de software desarrollado por los Archivos Federales de Suiza (PRONOM, 2010; Library of Congress, 2015). Un fichero SIARD consiste en un contenedor ZIP64, que alberga una colección de ficheros XML/XSD: uno por cada tabla (que captura su contenido), y uno general que contiene metadatos que definen la estructura de la base de datos, así como aspectos de autenticidad e integridad de datos. Posiblemente sea el formato de conservación de bases de datos más extendido. Ha sido adoptado como estándar (CH0165) en Suiza (Swiss Federal Archives, 2013). SIARD se basa en cuatro estándares internacionales XML, SQL:2008, UNICODE and ZIP64. La aplicación SIARD SUITE (Swiss Federal Archives, 2020) puede usarse para conservar bases de datos relacionales comerciales como Oracle, Microsoft SQL Server, MySQL, DB/2 y Microsoft Access
SIARD 2.0	.siard	Como parte del proyecto E-ARK (E-ARK, 2017a), se evolucionó el formato SIARD 1.0, resultando el nuevo formato SIARD 2.0 (E-ARK, 2017b). Las mejoras de la versión 2.0 incluyen soporte para SQL, mejores reglas de validación de datos, posibilidad de almacenar objetos de gran tamaño fuera del propio fichero SIARD y cambios en el mecanismo de compresión de ficheros
SIARD 2.2	.siard	En general, en SIARD los objetos de gran tamaño (conocidos como "LOB" o large objects) pueden guardarse: internamente al sistema de gestión de base de datos, o fuera del mismo. La versión SIARD 2.2 aporta las funcionalidades necesarias para asegurar la escalabilidad: en especial, el almacenamiento de LOB, cuando éste se hace de manera externa al fichero SIARD y fuera de la base de datos, se implementa mediante la extensión SQL/MED del estándar SQL:2008. En SIARD 2.0, el soporte al almacenamiento de objetos externos tanto al fichero SIARD como a la base de datos se hacía exclusivamente mediante enlaces basados en las URI de tipo "file:", mientras que SIARD 2.2 permite que estos LOB estén segmentados en varias carpetas y trozos de fichero, en vez de un fichero único.

Formato de fichero	Extensión	Observaciones
SQLite	.sqlite, .sqlite3, .db	SQLite es un formato *ligero* de fichero que permite almacenar una base de datos en un solo fichero (que incluye el tanto el motor de la base de datos, como éstos, divididos en "páginas") denominado fichero de base de datos principal. Este fichero se apoya en un segundo fichero temporal (*rollback journal*) que permitiría devolver la consistencia al fichero principal en caso de fallo durante una "transacción". La simplicidad de este formato permite que se puede usar tanto como fichero de respaldo o de almacenamiento. - Cabe reseñar que SQLite presenta **limitaciones** o restricciones, principalmente en cuanto a: tamaño máximo de almacenamiento de la base de datos, tráfico que puede soportar (estando orientada a aplicaciones de tráfico bajo-medio), limitaciones de la longitud máxima de las consultas y del número de columnas en las tablas, concurrencia de conexiones (ya que sólo permite trabajar en modo monousuario), y que sólo permite trabajar con las vistas de base de datos en modo lectura. - Las principales **ventajas** de SQLite son: su ligereza (que la hace adecuada para su uso en dispositivos embebidos como televisiones, cámaras, teléfonos móviles, etc.), su elevado rendimiento para operaciones de lectura y escritura, su portabilidad a todo tipo de entornos no requiere instalación ni configuración, y va asociada a un coste y complejidad reducidos (siendo una aplicación de software libre).

99. En la tabla anterior se hace alusión al uso de ficheros con extensión *.TXT* que contengan valores separados por un delimitador. También puede darse el caso de tener bases de datos, de cualquier propósito, o tecnología, ya sean generales o específicas, cuyo propósito principal sea contener y dar acceso a información en texto plano sin estructurar.

7.1. Recomendaciones sobre formatos para facilitar la explotación y reutilización

100. Se dan a continuación una serie de recomendaciones, dada la ausencia de contenido *semántico* en este tipo de formatos, para permitir una mejor explotación de la información y su reutilización. Se recomienda, por orden de prioridad, y según el estado de las citadas bases y su propósito de continuidad en el ámbito competente, las siguientes medidas (que no son exclusivas de las bases de datos de texto, sino aconsejables para todo tipo de bases de datos):

i. Categorización y metadatado de la información, siendo recomendable, en su caso, la lectura y la categorización automática, así como el metadatado si no fuera posible, o no se considerase conveniente, realizar estas operaciones por medios humanos. Caso de no poder hacer categorización y metadatado (ya sea automática o manual), al menos se recomienda hacer un indexado automático.

ii. Depuración y estructuración de los datos, del conjunto de datos de la actual base de datos, ya sea manual o automática, en función de la tecnología, volumetrías y recursos disponibles.

iii. Caso de que se desee migrar el contenido de dichas bases se recomienda la construcción de un modelo ontológico, sobre la base de una estructura o modelo de datos de destino definido previamente, y (en el caso de que la base de datos

contenga información textual) de los vocabularios normalizados que permitirán consolidar los procesos de lectura automática utilizando NLP-ML [Machine learning (ML) for natural language processing (NLP)].

iv. En el caso del punto 3 anterior, la categorización, metadatado, enriquecimiento y la consolidación de todo ello se articularían mediante un grafo de conocimiento de la base de datos.

v. Finalmente, en caso de los puntos 3 y 4 anteriores, se procedería a la construcción de los servicios de interrogación y descubrimiento de conocimiento que pertenecerían a la API final y podrían, en su caso, dar lugar a servicios de publicación interna y/o externa.

7.2. Caracterización de formatos

101. A la hora de identificar, automatizadamente, los formatos de fichero de objetos digitales se puede usar una herramienta como *DROID* (The National Archives, n.d.), *FIDO (Format Identification for Digital Objects)* (Open Preservation Foundation, 2020), o *Siegfried* (https://github.com/richardlehane/siegfried Lehane, 2020) que usa como fuente el registro de formatos de ficheros PRONOM (The National Archives, 2020).

102. Si se usa el formato SIARD para la conservación, hay que tener en cuenta que existen herramientas como el Kit de herramientas SIARD para la conservación de bases de datos (DBPTK) (KEEP, 2020) que puede realizar automáticamente la validación de la base de datos. (RDB SIARD, 2019).

8. HERRAMIENTAS DISPONIBLES PARA CONSERVACIÓN Y TRATAMIENTO DE BB.DD.

8.1. Exportación de bases de datos relacionales

103. Existen diversas herramientas, de código abierto o comerciales, que permiten la exportación de bases de datos relacionales a formato *SIARD* u otros formatos abiertos, y su posterior interpretación o tratamiento, entre las que cabe destacar:

i. DBPTK (Database Preservation Toolkit): herramienta de código abierto, que ofrece la posibilidad de exportar a varios formatos SIARD (SIARD 1, SIARD 2, SIARD DK); permite añadir en los metadatos de SIARD descripciones relativas a la base de datos, sus filas o columnas para facilitar la comprensión; tiene un módulo para la realización de búsquedas, y dispone de un módulo SIARD VALIDATION para validar los ficheros SIARD al finalizar el proceso de archivado. DBPTK posee una opción de Gestión (Manage) que permite administrar y filtrar los datos del fichero SIARD antes de cargarlo para ver los datos, y esta opción es muy útil, ya que permite que los ficheros se carguen más rápido una vez filtramos algunas tablas.

ii. SIARD Suite: aplicación de código abierto, desarrollada por los archivos federales suizos, que extrae contenido de bases de datos relacionales y lo almacena en el formato SIARD, como un conjunto de ficheros XML (que contienen el esquema de la base de datos, documentación de tablas y campos, y los datos propiamente dichos). De esta manera, se asegura la conservación a largo plazo de la base de datos. La interfaz gráfica posee un buscador para filtrar datos.

iii. FULL CONVERT: aplicación comercial para la conversión o migración de bases de datos, con un gran catálogo de más de 40 bases de datos soportadas, y que permite también la exportación a formato SIARD, y su importación sobre otro sistema de bases de datos, pudiendo emplear un planificador que realice esta tarea durante la noche. Permite filtrar por esquema y tablas, así como consultar los datos, ante de proceder con la conversión a SIARD. A diferencia de las anteriores herramientas, no dispone de interfaz de línea de comandos.

104. Un detalle para considerar, al escoger una aplicación para conservación en SIARD, es que cada una soporta un conjunto determinado de sistemas de bases de datos y de sistemas operativos y versiones compatibles. Además, las diferentes herramientas pueden presentar diferencias de rendimiento en cuanto a: la generación del backup en SIARD, la consulta, y la importación de datos. Asimismo, pueden disponer de interfaces gráficas o de línea de comandos de mayor o menor potencia y expresividad, y ofrecer diversas utilidades para el usuario final.

8.2. Herramientas de procesamiento del lenguaje natural para caracterizar las BB.DD.

105. Las tecnologías de la Web Semántica (con estándares como OWL, RDF, RDFa, SKOS, SPARQL y otros), estándares de metadatos (Dublin Core, DCAT y los estándares recogidos en el ENI) y el *Procesamiento del Lenguaje Natural* sobre texto libre (PLN), tienen como objetivo comprender el lenguaje, especialmente aquel que está depositado en los textos. «Comprender el lenguaje» significa en el ámbito de la Inteligencia Artificial hacerlo interrogable tanto para máquinas como para personas.

106. Los *Grafos de Conocimiento*, el resultado explotable de la aplicación de las tecnologías semánticas, y el PLN (*Machine Learning, Deep Learning*) están en un proceso de convergencia con el fin de desarrollar productos de *Inteligencia Artificial* más robustos, conversacionales y, en definitiva, más próximos al modo en el que las personas pensamos y nos relacionamos con la información y el conocimiento. Las tecnologías de Web Semántica, los Grafos de Conocimiento y el PLN pueden interactuar, mejorarse y ayudarse entre sí.

107. Bajo la denominación *Named Entity Recognition and Disambiguation (*NERD) se incluye una colección de APIs [Application Programming Interface] que permiten el procesamiento, comprensión, interpretación y análisis del lenguaje natural (*Natural Language Understanding* -NLU) orientadas al reconocimiento, extracción y desambiguación de entidades, conceptos y palabras clave en textos estructurados y no estructurados.

108. En esta nueva IA pueden y deben converger algunas de las técnicas que se han mostrado particularmente útiles en tareas concretas y que pueden resultar de ayuda cuando el volumen de los conjuntos de datos hace casi imposible que sean gestionados por humanos. Se están aplicando modelos de redes neuronales u otras técnicas de *aprendizaje automático* en tareas de clasificación de la información que forman parte de los *procesos ETL* (*Extract Transform Load*) hacia el grafo de conocimiento, como la extracción de tópicos, la categorización de entidades o el análisis de similitud semántica de textos; y otras técnicas y algoritmos sobre los grafos ya construidos, en tareas como el análisis de similitud de grafos o subgrafos o el reconocimiento y desambiguación de entidades (NERD).

109. La codificación y la construcción de estructuras de relaciones puede ser una base para la posterior construcción del modelo semántico.

8.3. Procedimientos de conservación

8.3.1. Emulación

110. Se puede optar por aplicar una **emulación**. “La emulación implica la sustitución de componentes software y/o hardware del sistema de base de datos por software que simula la forma de operar de esas partes; por ejemplo, el uso de una máquina virtual que imita el hardware manteniendo el resto de la pila tecnológica” (*Preserving Transactional Data,* Sarah Day Thompson).

111. Las ventajas principales del procedimiento de Emulación serían: la fidelidad en la simulación o imitación por software de la operativa del entorno original, y que no es necesario mantener entornos de hardware antiguos que pueden estar obsoletos o en desuso. Entre las desventajas de esta aproximación podemos citar: la dificultad de mantener estos entornos de emulación por software, que son complicados tecnológicamente y por tanto habitualmente caros, la dificultad de configuración, que se pueden presentar problemas inherentes a la tecnología original (por ejemplo, riesgos de seguridad característicos de los sistemas antiguos), y la necesidad de que los usuarios conozcan la operativa para trabajar con sistemas obsoletos.

8.3.2. Migración

112. La siguiente opción es una **migración** para convertir los datos a

i. otro formato de base de datos,
ii. un formato de conservación como SIARD (SFA SIARD, 2020),
iii. o archivos planos tipo ASCII o UTF-16 para texto, XML para datos y metadatos, y SQL estándar para la estructura de la base de datos

113. Las ventajas principales del procedimiento de Migración serían: la facilidad para reutilizar la información, y no requiere mantener hardware ni software específicos. Como desventajas de este enfoque, destacaríamos: la posible pérdida de datos durante la conversión (que podría mitigarse mediante procedimientos de Aseguramiento de la Calidad de Datos), la posible necesidad de migrar de nuevo en el futuro en función de las evoluciones tecnológicas, y que la migración de datos debe ir acompañada de otras medidas para preservar las características del entorno y contexto en el que opera la base de datos, lo cual puede entrañar complejidad.

8.3.3. Copia

114. Finalmente, se puede optar por obtener una **copia** estática (Microsoft, 2016), para conservar los datos. Esta opción hace que se siga dependiendo del Sistema de gestión de bases de datos original (DBMS).

115. La conservación también implica llevar a cabo las migraciones de soporte de almacenamiento.

116. o necesarias antes del fin de su vida útil. Asimismo, implica la realización de copias de seguridad periódicas de las bases de datos.

117. La Norma ISO 13008:2022 [Proceso de migración y conversión de documentos electrónicos] está pendiente de traducción, siendo la última versión en español la UNE-ISO 13008:2013. El estándar contempla los aspectos de planificación, requisitos y procedimientos para la conversión y/o migración de los documentos digitales para preservar su autenticidad, confiabilidad, integridad y usabilidad de dichos documentos como evidencia de funciones, actividades y trámites.

118. No obstante, la norma NO cubre de manera extensa los siguientes aspectos:

    i. Los sistemas de respaldo
    ii. La conservación de los documentos digitales
    iii. La funcionalidad de los repositorios digitales de confianza
    iv. El proceso de conversión los formatos físicos a digitales (y viceversa).

119. La norma pues, ayuda a la confección de un plan de conversión y/o migración con las mayores garantías posibles para seguir cumpliendo con el marco organizativo de la organización y los requisitos de gestión documental de dicha organización.

9. RECOMENDACIONES PARA LOS ORGANISMOS PRODUCTORES/GESTORES

9.1. Calendario de conservación y obligaciones legales

i. A la hora de considerar si se conserva una base de datos hay que considerar varios factores:

ii. Como y cuando se crearon y usaron los datos por un organismo

iii. El valor de los datos, especialmente si la base de datos está inventariada como de tipo “patrimonial” o valorada para su conservación permanente.

iv. Las obligaciones legales que condicionan los calendarios de conservación

v. El uso futuro que le darán otros usuarios.

120. El conocimiento de todas las regulaciones que afectan a los datos contenidos en la base de datos permitirá justificar su conservación.

121. Si se considera que debe conservarse la base de datos, será necesaria la coordinación entre los creadores de la base de datos, sus gestores o usuarios, personal TIC y todos aquellos relacionados con la gestión de la seguridad de la información y la protección de datos. Todos los actores implicados deben comprender los ***valores*** de la base de datos y las necesidades para su conservación.

122. En todo caso, para evitar que la obsolescencia de una base de datos pueda dar lugar la pérdida de información relevante, la valoración deberá realizarse durante la fase de explotación o vida útil de una base de datos, antes de que caiga en desuso.

123. En una organización hay que plantearse la creación de una guía o materiales de difusión y de capacitación para que los creadores, gestores y/o usuarios de las bases de datos sean conscientes de lo que deben hacer, durante la vida de la base de datos, para minimizar los riesgos de fallo en una transferencia y mitigar algunos de los riesgos de conservación a largo plazo.

124. Dentro de todos los procesos asociados a la conservación de las bases de datos, es conveniente conocer la estructura de la base de datos, controlando – si los hubiera – todos los documentos que recojan el control de cambios en dicha estructura, además de los campos de cada tabla (diferenciando, si fuera preciso entre el nombre real del campo y la etiqueta visible) y el tipo de campo de que se trata (numérico, alfanumérico, lógico, memo, etc.), con el número de caracteres de que consta, la máscara de entrada de datos (si la hubiera) y reglas de validación para cada campo; así como los valores predeterminados o las listas de origen de datos (si ciertos valores se extraen de lenguajes controlados). (RDB SIARD, 2019).

125. Hay que decidir cuándo y cómo capturar los datos. En unos casos los datos “activos” pueden ser capturados (y transferidos) varias veces durante la actividad de la base de datos. En otros sólo se transferirán datos inactivos o “históricos”. Estas decisiones deben consignarse en los calendarios de conservación y eliminación, buenas prácticas TIC y procedimientos de conservación y transferencia.

126. Es importante determinar qué contenido de información de una base de datos, así como qué funcionalidad proporcionada por el DBMS, la capa de aplicación, o el front-end de

usuario, es necesario/deseable conservar para satisfacer las necesidades de los usuarios y gestores.

127. Esto incluye usuarios, arquitectura y modelo de datos, así como el marco regulatorio de la gestión que pueda afectar a los calendarios de conservación, la protección de datos personales o cualquier restricción de tipo legal al uso y difusión de los datos.

128. También hay que tener en cuenta las convenciones sobre la nomenclatura de ficheros al guardar los datos exportados. Los nombres de fichero deben reflejar el nombre original de la base de datos, el nombre de la(s) tabla(s) que se exporta(n), la fecha de la exportación/importación o captura (Archaeology Data Service (ADS), 2009) y el nombre de la aplicación con la que se exporta o captura

129. Si una base de datos incluyera vínculos a recursos almacenados externamente se debiera considerar transferir dichos recursos al mismo tiempo que la base de datos.

9.2. Garantía de la Calidad de los datos

130. Un procedimiento recomendable es cotejar los datos de las bases de datos que se transfieran, comprobando que las estructuras de datos, tablas, relaciones y tipos de campo se corresponden con el modelo de datos y arquitectura declarados en la documentación de la base de datos

131. También se deben hacer comprobaciones en la estructura y el formato (ya sean tablas u hojas de cálculo, además de las fórmulas que se apliquen, consultas), comentarios o notas, datos ocultos o protegidos, caracteres especiales o delimitadores de campo y vínculos (ADS, 2009). Estas comprobaciones deben aplicarse tanto en una transferencia como en la migración de una base de datos a otro formato.

132. Es muy recomendable consultar con los creadores de los datos consignados en la base. También cotejar las diferencias que pudiera haber entre una copia de conservación y la base de datos activa. Si las discrepancias aparecieran durante la exportación, posiblemente se deba analizar el procedimiento para corregirlo y volverlo a lanzar sin que arroje fallos.

133. Antes de realizar la transferencia a archivo del contenido de una base de datos, puede ser aconsejable realizar un proceso de “Limpieza y depuración de datos”, encaminado al descubrimiento y corrección o eliminación de registros de datos erróneos, incompletos, inconsistentes o inutilizables de una tabla o base de datos. En un proceso de este tipo se podrían identificar, y posteriormente modificar o eliminar “datos sucios”, tales como: tipos incorrectos o fuera de rango, datos huérfanos, datos duplicados, datos mutuamente inconsistentes, categorización incorrecta, datos temporal o espacialmente incorrectos, datos desactualizados, erratas, ambigüedades, discrepancias en la codificación o formato de representación de la información, etc.

134. En la medida de lo posible, se debe acompañar el archivado de las bases de datos con toda la documentación necesaria para que permita su uso correcto en caso de que se decida su restauración, conteniendo toda la información disponible para ayudar a evitar sesgos indeseados, especialmente al emplear técnicas de *machine learning*.

135. La **documentación** de las bases de datos debe comprender aspectos tales como: la motivación por la que fueron creadas, el contexto de su creación y toda información sobre su procedencia; la composición y los problemas de calidad que puedan tener; indicar si es posible identificar individuos o subpoblaciones directa o indirectamente o si los datos

se refieren de alguna forma a personas con información sensible; si las bases de datos son una muestra de un conjunto más grande, indicar el procedimiento de muestreo; si existe algún informe de calidad o análisis de impacto del uso o si existen tareas para las que no debieran ser usadas.

i. toda la documentación necesaria (deseablemente tanto de la base de datos como de la capa de aplicación o incluso de la de presentación) para comprender los datos que contiene la copia de conservación y el contexto en el que fueron originados,

ii. y en caso de ser necesario, poner a disposición las herramientas de transformación de formatos de archivo a formatos de bases de datos, y herramientas que garanticen la posibilidad de visualización de los datos.

9.3. Documentación y metadatos

136. Es recomendable documentar las bases de datos de forma tan rica como sea posible. En línea con las recomendaciones de la Comisión Europea, se apuesta por el uso del estándar de metadatos DCAT, y que los metadatos incluyan al menos las propiedades obligatorias y, preferiblemente también las propiedades recomendadas.

137. Dentro de las actividades de gestión de la conservación de las bases de datos es necesario que se aborde de forma correcta la gestión de los metadatos relacionados con las bases de datos a preservar y la documentación asociada con el fin de conseguir los siguientes objetivos:

i. Metadatos efectivos y de calidad para mejorar el descubrimiento de las bases de datos preservadas facilitando su localización a través de catálogos, y otorgando además información sobre la calidad e integridad de los datos tras el archivado. Los metadatos se clasifican de acuerdo con el *marco de referencia* DAMA y la *Especificación UNE 0078:2023 Gestión del dato,* en tres tipos: de negocio, operativos y tecnológicos, los cuales se definen a continuación:

   a) **Metadatos de negocio**. Se enfocan en el contenido y en la condición o estado de los datos o bases de datos, e incluye detalles relacionados con su gobierno. Incluyen entre otros, nombres y definiciones de conceptos, áreas temáticas o entidades y atributos.
   b) **Metadatos técnicos**. Proporcionan información sobre los detalles técnicos de los de los datos y las bases de datos, sobre los sistemas que almacenan los datos o bases de datos y sobre los procesos que los mueven dentro y entre los sistemas.
   c) **Metadatos operativos**. Describen detalles del procesamiento y acceso a los datos o bases de datos. Incluyen información como permisos y reglas de compartición, registros de ejecución y extractos, o reglas de retención y archivado de datos.

ii. Entender mejor un ecosistema particular. Catalogar las bases de datos permite clasificar los ecosistemas de datos sectoriales, así como los actores que potencialmente pueden beneficiarse de dichas bases de datos preservadas.

iii. Resolver las necesidades de documentación para entender el contexto que facilite la conservación y, documentación sobre los aspectos éticos para prevenir el mal uso no intencionado en caso de que se decida la restauración de las bases de

datos y sirvan de fuente de información tanto para la investigación como para la alimentación de procesos de entrenamiento mediante técnicas de Inteligencia Artificial.

9.3.1. Gestión de los metadatos

138. Para llevar a cabo una correcta gestión de los metadatos de las bases de datos preservadas es necesario que se lleven a cabo las siguientes tareas, según se detalla en la Especificación UNE 0078:2023 Gestión del dato:

i. Definir los mecanismos que dan soporte a la gestión del metadato, lo que implica:

a) Una vez se han identificado las bases de datos que requieren conservación, se recolectan los metadatos existentes para completarlas, adecuarlas y migrarlas a estándares de metadatos que garanticen la interoperabilidad, tales como ISO/IEC 11179[10], Dublin Core, DCAT, o DCAT-AP11, siendo DCAT-AP la opción preferencial ya que es el vocabulario estándar desarrollado bajo la colaboración de la Comisión Europea para portales de datos en Europa, de tal forma que se busque la máxima compatibilidad y sinergia. En el documento referenciado al DCAT-AP, se muestra en detalle aquellas propiedades obligatorias, recomendables, u opcionales para cada una de las clases del perfil de aplicación ya sean estas propias o heredadas.

   Llegado el momento de documentar los metadatos relativos a la base de datos a preservar, se deberá rellenar un formulario o bien de forma manual o bien de forma automatizada, con los metadatos obligatorios, recomendados u opcionales según proceda para cada una de las clases del perfil de aplicación DCAT-AP ya sean estas propias o heredadas. Este formulario debe ser gestionado desde el archivo por herramientas de gestión de metadatos como CKAN.

b) Identificar las clasificaciones de las bases de datos (por tema, palabras clave, organismo que lo gestiona, etc.) que darán lugar a la creación de catálogos de metadatos. Se deben usar preferentemente listas controladas aceptadas internacionalmente o por la Unión Europea, según el área temática de la base de datos.

c) Determinar criterios de calidad del dato y de las bases de datos para el metadato y establecer relaciones entre los metadatos

d) Establecer mecanismos tecnológicos de compartición del metadato e identificar qué permisos deben ser otorgados a los diferentes roles para las operaciones de producción y consumición de metadatos.

ii. Crear o adaptar e integrar un soporte tecnológico para la gestión de repositorios del metadato, para lo que se debe:

---

[10] **ISO/IEC 11179-1:2023 INFORMATION TECHNOLOGY — METADATA REGISTRIES (MDR) — PART 1: FRAMEWORK. INFORMACIÓN ACCESIBLE EN** HTTPS://WWW.ISO.ORG/STANDARD/78914.HTML **EN ESTE ESTÁNDAR CUANDO SE HABLA DE METADATOS SE DEBE ENTENDER EL TÉRMINO COMO "DESCRIPCIÓN DE DATOS" NO INCLUYENDO UN TRATAMIENTO DE LOS METADATOS COMO TALES.**

[11] *https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/dcat-application-profile-data-portals-europe/release/211*

a) Desarrollar la infraestructura tecnológica necesaria para dar soporte tecnológico a las operaciones del metadato y adaptar la infraestructura tecnológica para integrar el soporte tecnológico a las operaciones de creación, lectura, actualización y borrado del metadato.

b) Desarrollar interfaces para la entrega del metadato que permita conectarlos con otras aplicaciones de gestión del dato e Implementar e integrar los controles necesarios para garantizar la calidad del metadato

iii. Poblar los repositorios del metadato, esto implica:

a) Para las bases de datos a preservar, extraer, priorizar, consensuar y almacenar los metadatos de negocio que pueden ser importantes o relevantes para comprender el contexto del negocio sobre esa base y sus datos.

b) Identificar, documentar y almacenar los metadatos operativos que permitan comprender el contexto operacional de las bases, así como los metadatos tecnológicos que permitan comprender el contexto tecnológico de las mismas.

c) Establecer las relaciones entre los metadatos, de forma que se complementen y permitan proporcionar una visión global del conocimiento que se tiene sobre las bases y representarlo a través de un grafo de conocimiento.

d) Verificar y validar de forma regular la calidad del metadato, y propagar todos los posibles cambios que se pudieran producir

iv. Publicar los metadatos y fomentar su uso, lo que implica:

a) Anunciar y comunicar a los principales interesados la disponibilidad de los metadatos en el repositorio de datos, animando al uso de los diferentes tipos de metadatos para un mejor soporte en el resto de los diferentes procesos de gobierno y gestión del dato.

b) Los metadatos asociados a las bases de datos preservadas deberán publicarse en el catálogo de datos de la AGE, que se hará disponible a través de la plataforma del dato de la SGAD (Secretaría General de Administración Digital)

9.3.2. Metadatos de conservación dentro del modelo OAIS

139. OAIS (Open Archival Information System) es uno de los *modelos de referencia* impulsados por el *Comité Consultivo para sistemas de datos del Espacio* (Consultative Committee for Space Data Systems; CCSDS). La “práctica recomendada” CCSDS 650.0-M-2 Reference Model for an Open Archival Information System (OAIS), se convirtió en el *estándar* ISO 14721 en 2002, publicándose su segunda versión en 2012[12]. Es un **modelo** de alto nivel (aunque **no** se pueda usar como un *modelo de implantación*) que describe los componentes y procesos necesarios para archivos digitales, distinguiendo seis áreas funcionales (*ingesta*, *archivado*, *gestión de datos*, *administración*, *planificación de la conservación* y *acceso*).

140. Al ser un *modelo*, y no entrar en la implantación, es neutral en lo que respecta a tipos de objetos y tecnologías usadas[13], existiendo múltiples herramientas para su implantación.

---

[12] *Disponible en la Biblioteca Electrónica del Ministerio de Hacienda a través de su intranet en la URL http://biblioteca.minhac.age/Home/Contenido?menu=14 que permite el acceso a la suscripción institucional a la consulta de bases de datos, entre las cuales están las de AENOR.*

[13] *En relación con el modelo de referencia se han desarrollado (y se deben desarrollar) una serie de estándares:*

Es, pues, un marco para la definición de una estrategia de conservación universal y definición de modelos propios enfocados en un área en particular propia para cada organización y para todo tipo de soportes digitales.

141. Los metadatos de conservación son los datos necesarios para cumplir con los requerimientos que permitan garantizar el uso a largo plazo de un recurso digital. En el contexto de metadatos de conservación, el paquete de información relevante es el AIP (*Archival Information Package* o Paquete de Información de Archivo), ya que este es el paquete que se conserva a largo plazo.

142. Es necesario distinguir entre el paquete de información que es preservado por un modelo OAIS y los paquetes de información que se envían o difunden desde el mismo. Estas variantes se denominan paquete de información de envío (SIP), paquete de información de archivo (AIP) y paquete de información de difusión (DIP).

    i. El paquete de información de envío (SIP) es ese paquete que un productor envía a un OAIS. Su forma y contenido detallado generalmente se negocian entre el productor y el OAIS. La mayoría de los SIP tendrán **información de contenido** (CI) e **información de descripción de la conservación** (PDI), conceptos que se describirán de manera más extensa en el próximo párrafo.

    ii. Las relaciones entre SIP y AIP pueden ser complejas; además de una relación uno a uno simple en la que un SIP produce un AIP, otras posibilidades incluyen:

        a) un AIP producido a partir de múltiples SIP producidos en diferentes momentos por un Productor o por muchos Productores;

        b) un SIP que da como resultado varios AIP; y muchos SIP de una o más fuentes que se desagregan y se recombinan de diferentes maneras para producir muchos AIP.

Incluso en el primer caso, el OAIS puede tener que realizar transformaciones en el SIP. La información de empaquetamiento siempre estará presente de alguna forma. Dentro del OAIS, uno o más SIP se transforman en uno o más paquetes de información de archivo (AIP) para su conservación.

- *Estándar para la metodología de envío (ingesta) a los archivos, ISO 20652:2006 (CCSDS 651.0-B-1:2004) Space data and information transfer systems—Producer-Archive Interface—Methodology Abstract Standard (PAIMAS);*
- *Estándar para el envío (ingesta) de fuentes de datos a archivos, ISO 20104:2015 (CCSDS 651.1-B-1) Space data and information transfer systems—Producer-Archive Interface Specification (PAIS);*
- *PREMIS Data Dictionary for Preservation Metadata. Version 3.0, PREMIS Editorial Committee, June 2015;*
- UNE-ISO 23081-1:2018 *Información y documentación. Procesos de gestión de documentos. Metadatos para la gestión de documentos. Parte 1: Principios.*
- UNE-ISO 23081-2:2021 *Información y documentación. Procesos de gestión de documentos. Metadatos para la gestión de documentos. Parte 2: Elementos conceptuales y de implementación.*
- UNE-ISO/TR 23081-3:2012 IN *Información y documentación. Metadatos para la gestión de documentos. Parte 3: Método de auto-evaluación.*
- ISO 16363:2012 (CCSDS 652.0-M-1) Space data and information transfer systems — Audit and certification of trustworthy digital repositories;
- ISO 16919:2014 (CCSDS 652.1-M.2) Space data and information transfer systems — Requirements for bodies providing audit and certification of candidate trustworthy digital repositories;

SIP — Paquete de sumisión de información
AIP — Paquete de archivo de información
DIP — Paquete de diseminación de información

143. Un AIP es la agregación de cuatro tipos de objetos de información. Cada uno de estos tipos se describe a continuación. Cabe destacar que cada objeto de información consiste en un Objeto de Datos y la Información de Representación asociada necesaria para darle sentido al Objeto de Datos.

i. **Información de Contenido** (*Content Information* o CI): consiste en el contenido del objeto de datos o *Content Data Object*, es decir, la información que debe conservarse, junto con su Información de Representación asociada.

ii. **Información de Descripción de la Conservación** (*Preservation Description Information* o PDI: contiene información necesaria para gestionar la conservación de la Información de Contenido o CI con la que está asociada. El modelo de referencia OAIS identifica cuatro tipos de PDI:

a) **Información de referencia** (*Reference Information*): Esta categoría se refiere a los identificadores únicos o códigos asignados a los objetos digitales para identificarlos tanto interna como externamente, para identificarlos de manera unívoca. Es importante destacar que, aunque la Información de referencia en el modelo OAIS incluye metadatos descriptivos, no se limita exclusivamente a ellos. También puede incluir otros tipos de identificadores o códigos necesarios para la gestión y la conservación de los objetos digitales.

b) **Información de contexto** (*Context Information*): Esta categoría documenta la relación de los objetos digitales con su entorno, como el propósito de su creación, relaciones con otros objetos o documentación relacionada, etc.

c) **Información de procedencia** (*Provenance Information*): Esta categoría documenta la historia de los objetos digitales, incluyendo su origen, historia de custodia y las intervenciones realizadas para garantizar su conservación, integridad y autenticidad. También incluye información sobre la cadena de custodia.

d) **Información de integridad** (*Fixity Information*): Esta categoría se refiere a los mecanismos utilizados para garantizar la integridad y autenticidad de los objetos digitales, como comprobaciones de integridad, firmas digitales, etc.

iii. **Información de Empaquetado** (*Packaging Information* o PI): vincula el objeto digital y sus metadatos asociados en una unidad o paquete identificable (es decir, un AIP o Paquete de Información de Archivo).

iv. **Información Descriptiva** (Descriptive Information o DI): facilita el acceso a la Información de Contenido o CI a través de las herramientas de búsqueda y recuperación del archivo. La Información Descriptiva sirve como entrada para los instrumentos de búsqueda del archivo y generalmente se deriva de la Información de Contenido o la Información de Descripción de Conservación. Como se ha mencionado anteriormente, es importante que esta información descriptiva (DI) pueda ser publicada en formato DCAT-AP para poder usarse en catálogos de nivel superior cuando proceda.

v. La AIP también puede contener una colección de otros AIP.

vi. La información de empaquetado del AIP se ajustará a los estándares internos de OAIS y puede variar según sea administrada por este último.

144. En respuesta a una solicitud, el OAIS proporciona todo o parte de un AIP a un consumidor en forma de paquete de información de difusión (DIP). El DIP también puede incluir colecciones de AIP y puede tener o no un PDI completo. La información de empaquetado (PI) estará presente para que el consumidor pueda distinguir claramente la información solicitada.

145. Como fruto del grupo trabajo PREMIS (Preservation Metadata: Implementation Strategies) se desarrolló un conjunto básico de metadatos de preservación aplicable a toda la casuística de trabajo que implica la conservación digital. Dicho conjunto debía acompañarse de instrucciones y recomendaciones para la creación, gestión y uso de metadatos, materializado en 2005 con la publicación del Informe Final del Grupo de Trabajo PREMIS: Diccionario de Datos para los metadatos de preservación. Además, publicaron un conjunto de schemas XML para apoyar la implantación del *Diccionario de Datos* en los sistemas de preservación digital.

146. El Diccionario de datos se basa el modelo de referencia OAIS (ISO 14721) que, como hemos visto, proporciona la base conceptual en forma de taxonomía de objetos de información y paquetes de archivo de objeto, así como la estructura de sus metadatos asociados. La versión 3.0 del Diccionario de Datos expande el alcance más allá de los límites de los repositorios para aplicarse a la representación de los metadatos durante todo el ciclo de vida de un objeto digital.

9.3.3. Documentación de las bases de datos para su conservación.

147. Adicionalmente a la creación de los metadatos, se pueden acompañar las BBDD con documentación que facilite la conservación y la futura restauración de las bases de datos. Para ello se pueden tomar en cuenta los siguientes aspectos:

9.3.3.1. Documentación del contexto

148. Según A. Millaruelo (2020), *para documentar el contexto, es necesario contar con una documentación adecuada de las características del sistema o sistemas de información que hacen uso de la base de datos, facilitaría la conservación no sólo de la capa de datos*

sino de las otras capas, especialmente de la de aplicación, evitando la pérdida de información que un proceso de conservación (por ejemplo, una migración) podría ocasionar. Entre esta documentación (que podría ir asociada a metadatos) encontraríamos:

i. Información de contexto, que incluya las reglas de negocio, las razones legales o técnicas que llevaron a la creación del sistema de información, etc.
ii. Manuales de usuario, capturas de pantalla, código de procedimientos, documentación de análisis o diseño, guías de instalación o mantenimiento, diagramas E/R (entidad/relación).
iii. Información técnica sobre el DBMS (sistema de gestión de bases de datos).
iv. Esquemas de la base de datos, de las tablas que la integran, de las relaciones entre ellas, de las restricciones, etc.
v. Determinación de los datos esenciales (por ejemplo, de las claves primarias y claves foráneas si es un DBMS relacional).
vi. Modelo de datos: define las relaciones entre entidades y la estructura de aquéllas.

9.3.3.2. Consideraciones éticas.

149. Según el trabajo de investigación *Datasheets for datasets*[14], se propone un marco de referencia para la documentación de las bases de datos que serán empleadas en el entrenamiento de modelos de *Machine Learning (ML),* el marco consiste en una serie de preguntas que guían la documentación de los datos incluyendo cuestiones acerca de todo su ciclo de vida.

150. El motivo para la creación de un *Datasheet for dataset* es que las características de los datos empleados en la formulación de los modelos de ML tienen una clara influencia en su comportamiento. Si los conjuntos de datos en los que se basa el modelo reflejan sesgos sociales indeseados, los modelos pueden reproducir o amplificar esos sesgos. Si bien es cierto, que estos procedimientos son más propios para bases de datos en activo, su uso sobre bases de datos preservadas o a preservar puede servir en caso de que los investigadores quieran restaurar las bases de datos y usarlas para el entrenamiento de modelos de ML u otros. Por estas razones se recomienda que en el momento en que se procede al archivado de una BBDD, se debería documentar la procedencia, creación y uso de los conjuntos de datos, con el fin de anticipar su posible uso con técnicas de ML, y de esta manera evitar resultados erróneos o discriminatorios. Para ello se propone la creación de una ficha que amplíe la documentación de la base de datos con información como la siguiente:

i. La motivación que originó la creación de la BBDD y el proceso de recolección, con información como quién creó o financió dicha base de datos, cómo se adquirieron los datos, o cómo fue el proceso de revisión ética. Se tratan especialmente los aspectos éticos del procesamiento de datos protegidos por el RGPD.
ii. La composición de la BBDD, sus unidades de observación (documentos, fotos, personas, lugares, etc.) y su procedencia (si ha sufrido transformaciones, si la base de datos es una muestra de otra más grande y en ese caso qué técnica de muestreo ha sido empleada, etc.).

---

[14] *Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, andKate Crawford. 2021. Datasheets for datasets. Commun. ACM 64, 12 (December 2021), 86–92.*

iii. Informes de calidad que detallen los procedimientos de preprocesamiento, limpieza o etiquetado. (p. ej., discretización o agrupamiento, tokenización[15] y segmentación en el procesamiento del lenguaje natural, para dividir un texto en unidades más pequeñas que constituyen las unidades para el procesamiento semántico; etiquetado de partes de un texto, extracción de funciones SIFT[16] y eliminación de instancias o imputación de valores).

iv. Descripción del uso que se ha dado en el pasado al contenido de la base, así como información relacionada con posibles riesgos de un mal uso futuro, de tal forma que se puedan prevenir en la mayor medida posibles daños, sesgos o tratos injustos.

[15] El término “tokenization” tiene al menos dos significados distintos en la informática, según se trate del campo de la seguridad informática y blockchain, o bien el procesamiento del lenguaje natural. En este documento aplica la segunda acepción.
[16] Las funciones SIFT (scale-invariant feature transform) sólo aplican al caso de objetos de tipo imagen.

10. DEFINICIONES Y ACRÓNIMOS

10.1. Definiciones

Acceso: Derecho de los ciudadanos a la consulta de la información pública de acuerdo con la normativa vigente, en aplicación del art. 105b de la Constitución Española.

Actuación administrativa automatizada: Actuación administrativa producida por un sistema de información adecuadamente programado sin necesidad de intervención de una persona física en cada caso singular. Incluye la producción de actos de trámite o resolutorios de procedimientos, así como de meros actos de comunicación.

Administrador del sistema/de la seguridad del sistema: persona encargada de la instalación y el mantenimiento de un sistema de información, implantando los procedimientos y la configuración de seguridad que se haya establecido en el marco de la política de seguridad del organismo.

Análisis de riesgos: estudio de las consecuencias previsibles de un posible incidente de seguridad, considerando su impacto en la organización (en la protección de sus activos, en su misión, en su imagen o reputación, o en sus funciones) y la probabilidad de que ocurra.

Anonimización: Proceso por el cual la información personal (IP) se altera de forma irreversible de tal manera que el titular de la IP ya no puede ser identificado directa o indirectamente, ya sea por el controlador de la PI solo o en colaboración con cualquier otra parte.

Aprendizaje automático (*Machile learning*): Proceso que utiliza algoritmos, en lugar de codificación procedimental, y que permite aprender de los datos existentes para predecir resultados futuros

Archivo de Seguridad/Respaldo: Sistema o depósito en el que se conservan las copias en cualquier soporte de documentos originales, conjuntos de datos u otros recursos de información, con la finalidad de garantizar la disponibilidad de aquéllos en caso de pérdida o destrucción por un incidente de seguridad.

Arquitectura: Conceptos o propiedades fundamentales de un sistema en su entorno, plasmados en sus elementos, relaciones y en los principios de su diseño y evolución

Arquitectura de seguridad: conjunto de elementos físicos y lógicos que forman parte de la arquitectura del sistema y cuyo objetivo es la protección de los activos dentro del sistema y en las interconexiones con otros sistemas.

Auditoría de la seguridad: es un proceso sistemático, independiente y documentado que persigue la obtención de evidencias y su evaluación objetiva para determinar en qué medida se cumplen los criterios de auditoría en relación con la idoneidad de los controles de seguridad adoptados, el cumplimiento de la política de seguridad, las normas y los procedimientos operativos establecidos, y detectando desviaciones a los antedichos criterios.

Autenticación: *ratificación de la identidad de un usuario, proceso o dispositivo.*

Autenticidad: Referido a un documento o conjunto de datos; propiedad que puede atribuírsele como consecuencia de que puede probarse que es lo que afirma ser, que ha sido creado por la persona o entidad de la cual se afirma que lo ha creado, y que ha sido creado en el momento en que se afirma, sin que haya sufrido ningún tipo de modificación.

Calidad del dato: Grado en que las características de los datos satisfacen necesidades declaradas e implícitas cuando se utilizan en condiciones especificadas.

Categoría de seguridad de un sistema: es un grado, dentro de la escala Básica-Media-Alta, con el que se adjetiva un sistema de información a fin de seleccionar las medidas de seguridad necesarias para el mismo. La categoría de seguridad del sistema recoge la visión holística del conjunto de activos como un todo armónico, orientado a la prestación de unos servicios.

Certificado de firma electrónica (factor de autenticación): una declaración electrónica que vincula los datos de validación de una firma con una persona física o jurídica y confirma, al menos, el nombre o el seudónimo de esa persona.

Certificado cualificado de firma electrónica: un certificado de firma electrónica que ha sido expedido por un prestador cualificado de servicios de confianza y que cumple los requisitos establecidos en el anexo I del Reglamento (UE) 910/2014, del Parlamento Europeo y del Consejo, de 23 de julio de 2014, relativo a la identificación electrónica y los servicios de confianza para las transacciones electrónicas en el mercado interior.

Ciberamenaza: *amenaza a los sistemas y servicios presentes en el ciberespacio o alcanzables a través de éste.*

Ciberataque: *cualquier conducta dolosa de individuos u organizaciones, conocidos o no, desarrollada a través del ciberespacio contra sistemas de información, con el propósito de sustraer, alterar, abusar, desestabilizar, inutilizar, destruir o eliminar activos.*

Ciberincidente: *Incidente relacionado con la seguridad de las tecnologías de la información y las comunicaciones que se produce en el ciberespacio.*

Ciberseguridad (seguridad de los sistemas de información): *la capacidad de las redes y sistemas de información de resistir, con un nivel determinado de fiabilidad, toda acción que comprometa la disponibilidad, autenticidad, integridad o confidencialidad de los datos almacenados, transmitidos o tratados, o los servicios correspondientes ofrecidos por tales redes y sistemas de información o accesibles a través de ellos.*

Ciclo de vida de un documento electrónico: Conjunto de las etapas o períodos por los que atraviesa la vida del documento, desde su identificación en un sistema de gestión de documentos, hasta su selección para conservación permanente, de acuerdo con la legislación sobre de aplicación en cada caso, o para su destrucción controlada previa autorización del organismo competente.

Compromiso de la seguridad: *incidente de seguridad en el que, debido a una violación de las medidas técnicas u organizativas de seguridad, una información o un servicio quedan expuestos, o potencialmente expuestos, a un acceso no autorizado.*

Conjunto de datos: *Agrupación de datos con sentido lógico usable para una tarea determinada*

Confidencialidad: propiedad o característica consistente en que la información ni se pone a disposición, ni se revela a individuos, entidades o procesos no autorizados.

Conjunto de datos de alto valor: Según la Directiva UE 2019/1024 son documentos cuya reutilización está asociada a considerables beneficios para la sociedad, el medio ambiente y la economía, en particular debido a su idoneidad para la creación de servicios de valor añadido, aplicaciones y puestos de trabajo nuevos, dignos y de calidad, y del número de beneficiarios potenciales de los servicios de valor añadido y aplicaciones basados en tales conjuntos de datos;

Conservación: Conjunto de procesos y operaciones dedicados a asegurar la permanencia intelectual y técnica de los activos de información y documentos administrativos a lo largo del tiempo.

Custodia: Responsabilidad jurídica que implica, por parte de una institución, organismo o persona, el control y la adecuada conservación de los activos de información y el patrimonio documental a su cargo, cualquiera que sea la titularidad de los mismos.

Dato: Una representación de hechos, conceptos o instrucciones de un modo formalizado, y adecuado para su comunicación, interpretación o procesamiento por medios automáticos o humanos.

Datos de carácter personal: Cualquier información numérica, alfabética, gráfica, fotográfica, acústica o de cualquier otro tipo concerniente a personas físicas identificadas o identificables. Se definen en el artículo 4, punto 1, del Reglamento (UE) 2016/679;

Datos maestros: Datos que posee una organización para describir las entidades que son independientes y fundamentales para dicha organización, ya las que se hace referencia para realizar sus transacciones**.**

Destrucción: Borrado lógico o *disposición* física, de todos los soportes de información que contengan documentos o información administrativa, o conjuntos de datos relacionados. Es el proceso final de la actividad de eliminación controlada de documentación/información que haya perdido su valor administrativo, probatorio o constitutivo o extintivo de derechos y que no haya desarrollado, ni se prevea que vayan a desarrollar, valores históricos. La destrucción se debe realizar por cualquier método que garantice la imposibilidad de reconstrucción de los documentos y tendrá que adecuarse al contenido de los soportes y la calificación de la documentación según afecte este contenido a datos personales, o la seguridad nacional.

Disponibilidad: Referido a un documento, indica propiedad o característica del mismo que permite que éste pueda ser consultado, localizado, recuperado, presentado o interpretado. El documento debe señalar la actividad o actuación donde se generó, proporcionar la información necesaria para la comprensión de las actuaciones que motivaron su creación y utilización, identificar el contexto marco de las actividades y las funciones de la organización y mantener los vínculos existentes con otros documentos como reflejo de una secuencia de actuaciones.

Dispositivo de autenticación (token): autenticador físico.

Documento: Información o contenido de cualquier naturaleza archivada en un soporte físico o electrónico y susceptible de identificación y tratamiento diferenciado.

> ***Documento electrónico:*** Información de cualquier naturaleza en forma electrónica, archivada en un soporte electrónico según un formato determinado y susceptible de identificación y tratamiento diferenciado.
>
> ***Documento administrativo electrónico:*** Objeto digital administrativo que contiene la información objeto (datos) y los datos asociados a ésta (firma y metadatos). En el marco del ENI, este concepto incluye tanto los documentos electrónicos producidos por las Administraciones públicas en el ejercicio de sus competencias como los documentos electrónicos aportados por los ciudadanos en el contexto de un procedimiento dado.

Elementos de datos: *Unidad de datos identificable más pequeña dentro de un contexto determinado cuya definición, identificación, valores permitidos y otra información se especifica mediante un conjunto de propiedades. También puede considerarse cualquier artefacto que esté involucrado en el uso, gestión y explotación de los datos.*

Eliminación: Procedimiento que consiste en la identificación de los documentos y recursos de información que se van a destruir, de manera controlada, conforme a los plazos establecidos en la fase de valoración.

Especificación de datos: Conjunto de requisitos que cuben las características de los datos aptos para uno o más fines particulares.

Esquema de metadatos: Instrumento que define la incorporación y gestión de los metadatos de contenido, contexto y estructura de los documentos electrónicos a lo largo de su ciclo de vida.

Estándar: Especificación técnica aprobada por un organismo de normalización reconocido para una aplicación repetida o continuada cuyo cumplimiento no sea obligatorio y que esté incluida en una de las categorías siguientes:

> **Norma europea:** norma adoptada por un organismo europeo de normalización y puesta a disposición del público.
>
> **Norma internacional:** norma adoptada por una organización internacional de normalización y puesta a disposición del público.
>
> **Norma nacional:** norma adoptada por un organismo nacional de normalización y puesta a disposición del público.

Expediente electrónico: Conjunto de documentos electrónicos correspondientes a un procedimiento administrativo, cualquiera que sea el tipo de información que contengan.

Expolio: Enajenación o destrucción ilegal de documentos o recursos de información que forman parte del patrimonio documental tanto por acción, como por omisión. Según la naturaleza de los documentos o recursos afectados puede suponer que se incurra en otras figuras penales como la infidelidad en la custodia de documentos públicos o la destrucción no autorizada de datos informáticos.

Fiabilidad: Referido a un documento, propiedad o característica que indica que su contenido puede ser considerado una representación completa y precisa de las actuaciones, las actividades o los hechos de los que da testimonio y al que se puede recurrir en el curso de posteriores actuaciones o actividades.

Fichero de datos: Conjunto de registros de datos relacionados que pueden ser tratados como una unidad.

Formato de datos: Disposición de los datos para su almacenamiento o visualización.

***Gestión de incidentes*:** procedimientos seguidos para detectar, analizar y limitar un incidente y responder ante éste.

Gestión de riesgos: actividades coordinadas para dirigir y controlar a una organización con respecto a los riesgos.

Gestión del dato: *Desarrollo, ejecución y supervisión de planes, políticas, programas y prácticas que proporcionen, controlen, protejan y mejoren el valor de los activos de datos e información a lo largo de su ciclo de vida.*

Gestión de documentos: Conjunto de operaciones dirigidas al control eficaz y sistemático de la creación, recepción, uso, valoración y conservación de los documentos, incluidos los procesos para incorporar y mantener pruebas de las actuaciones o actividades de dicha organización, en forma de documentos y sistemas de información.

Gestión de riesgos: actividades coordinadas para dirigir y controlar a una organización con respecto a los riesgos.

Gobierno del dato: Desarrollo y aplicación de políticas relacionadas con la gestión del dato.

Grafo[17]***:*** Diagrama que representa mediante puntos y líneas las relaciones entre pares de elementos y que se usa para resolver problemas lógicos, topológicos y de cálculo combinatorio.

Identidad: Conjunto de características de un documento, o activo de información, que lo identifican de manera única y lo distinguen de cualquier otro documento o activo. Junto con la integridad, un componente de la autenticidad.

Incidente de seguridad (ciberincidente o incidente): suceso inesperado o no deseado con consecuencias en detrimento de la seguridad de las redes y sistemas de información.

Información Pública: contenidos o documentos, cualquiera que sea su formato o soporte, que hayan sido elaborados o adquiridos en el ejercicio de sus funciones y que obren en poder de alguna de las administraciones públicas y otras entidades recogidas en el ámbito subjetivo de aplicación de la normativa sobre transparencia y acceso.

Integridad: propiedad o característica consistente en que el activo de información no ha sido alterado de manera no autorizada. La integridad es un componente de la autenticidad junto a la identidad.

Intercambio de datos: Almacenamiento, acceso, transferencia y archivo de datos.

Medidas de seguridad: conjunto de disposiciones encaminadas a proteger al sistema de información de los riesgos a los que estuviere sometido, con el fin de asegurar sus objetivos de seguridad. Puede tratarse de medidas de prevención, de disuasión, de protección, de detección y reacción, o de recuperación.

Metadato: Dato que define y describe otros datos. Existen diferentes tipos de metadatos según su aplicación.

Metadato de gestión de documentos: Información estructurada o semi-estructurada que hace posible la creación, gestión y uso de documentos a lo largo del tiempo en el contexto de su creación. Los metadatos de gestión de documentos sirven para identificar, autenticar y contextualizar documentos, y del mismo modo a las personas, los procesos y los sistemas que los crean, gestionan, mantienen y utilizan.

Modelo de datos: *Conjunto de definiciones (modelo conceptual), interrelaciones (modelo lógico) y reglas y convenciones (modelo físico) que permiten describir los datos para su intercambio*

Organismo Productor: Persona física o jurídica, pública o privada, que en el ejercicio de sus *competencias* genera un fondo documental, conjuntos de datos o activos de información asociados a su actividad.

---

[17] *https://dle.rae.es/grafo*

Patrimonio Documental: Totalidad de documentos de cualquier época generados, conservados o reunidos en el ejercicio de su función por cualquier organismo o entidad de carácter público, por las personas jurídicas en cuyo capital participe mayoritariamente el Estado u otras entidades públicas y por las privadas, físicas o jurídicas, gestoras de servicios públicos en lo relacionado con la gestión de dichos servicios. También los documentos con una antigüedad superior a los cuarenta años, generados conservados o reunidos en el ejercicio de sus actividades por las entidades y asociaciones de carácter político, sindical o religioso y por las entidades, fundaciones y asociaciones culturales y educativas de carácter privado. Por último, integran el Patrimonio Documental los documentos con una antigüedad superior a los cien años, generados, conservados o reunidos por cualesquiera otras entidades particulares o personas físicas.

Plazo preventivo de conservación: Período durante el cual se mantiene en suspenso la decisión de eliminar una serie documental o un recurso de información, a pesar de que hayan prescrito sus valores primarios, para prevenir la posibilidad de que pueda tener alguna utilidad administrativa.

Plazo de reserva: Período de tiempo durante el cual, según la normativa vigente, una documentación no puede ser consultada públicamente.

Plazo de vigencia administrativa: Período, reglado por las normas de procedimiento, durante el cual los valores primarios de una serie documental, de manera simultánea o sucesiva, mantienen su actividad en mayor o menor grado.

Prescripción: Momento en el que se extinguen los valores primarios de los documentos administrativos.

Política de gestión de documentos electrónicos: Orientaciones o directrices que define una organización para la creación y gestión de documentos auténticos, fiables y disponibles a lo largo del tiempo, de acuerdo con las funciones y actividades que le son propias. La política se aprueba al más alto nivel dentro de la organización, y asigna responsabilidades en cuanto a la coordinación, aplicación, supervisión y gestión del programa de tratamiento de los documentos a través de su ciclo de vida.

Procedimiento: *Medidas o dispositivos necesarios para lograr las directrices de las políticas y que pueden ser de carácter organizativo o técnico.*

Procedimiento administrativo: Proceso formal regulado jurídicamente para la toma de decisiones por parte de las Administraciones Públicas para garantizar la legalidad, eficacia, eficiencia, calidad, derechos e intereses presentes, que termina con una resolución en la que se recoge un acto administrativo; este proceso formal jurídicamente regulado se implementa en la práctica mediante un proceso operativo que coincide en mayor o menor medida con el formal.

Proceso: conjunto organizado de actividades que se llevan a cabo para producir un producto o prestar un servicio, que tiene un principio y fin delimitados, que implica recursos y da lugar a un resultado.

Proceso de seguridad: método que se sigue para alcanzar los objetivos de seguridad de la organización. El proceso se diseña para identificar, medir, gestionar y mantener bajo control los riesgos a que se enfrenta el sistema en materia de seguridad.

Proceso operativo: Conjunto organizado de actividades que se llevan a cabo para producir un producto o servicio; tiene un principio y fin delimitado, implica recursos y da lugar a un resultado.

Proceso TIC: conjunto de actividades llevadas a cabo para la concepción, elaboración, suministro y mantenimiento de un producto o servicio TIC.

Repositorio electrónico: Archivo centralizado donde se almacenan y administran datos y documentos electrónicos, y sus metadatos.

Requisitos mínimos de seguridad: exigencias mínimas necesarias para asegurar la información tratada y los servicios prestados.

Repositorio electrónico: Archivo centralizado donde se almacenan y administran datos y documentos electrónicos, y sus metadatos.

Responsabilidad: Combinación de actividades, toma de decisiones y consecución de resultados cuando la combinación es realizada por una parte determinada de la organización.

Reutilización: Según la Directiva UE 2019/1024, el uso por personas físicas o jurídicas de documentos que obran en poder de:

a) organismos del sector público, con fines comerciales o no comerciales distintos del propósito inicial que tenían esos documentos en la misión de servicio público para la que se produjeron, excepto para el intercambio de documentos entre organismos del sector público en el marco de sus actividades de servicio público, o

b) empresas públicas, con fines comerciales o no comerciales distintos del propósito inicial que tenían esos documentos de prestar servicios de interés general para el que se produjeron, excepto para el intercambio de documentos entre empresas públicas y organismos del sector público que se realice exclusivamente en el desarrollo de las actividades de servicio público de los organismos del sector público;

Rol: *Conjunto de responsabilidades que pueden ser asignadas por una organización a una persona y esta asignación dirige a la persona a realizar cada responsabilidad.*

Sistema de gestión de documentos**:** Marco definido por la política de gestión de documentos de una organización donde se integran los recursos y se implantan las prácticas de gestión de documentos (establecidas en forma de programa de tratamiento) en una organización. Integran el sistema, además de la política, los recursos, el programa de tratamiento y los documentos incorporados a dicho sistema.

Sistema de información: cualquiera de los elementos siguientes: *redes de comunicaciones* que utilice la entidad; *dispositivo o grupos de dispositivos* interconectados o relacionados entre sí, en el que uno o varios de ellos realicen, mediante un programa, el tratamiento automatizado de datos digitales; los *datos digitales* almacenados, tratados, recuperados o transmitidos mediante los elementos contemplados anteriormente, incluidos los necesarios para el funcionamiento, utilización, *protección y mantenimiento de dichos elementos.*

Soporte: *Objeto sobre el cual o en el cual es posible grabar y recuperar datos.*

Trámite: *Cada uno de los estados y diligencias que hay que recorrer en un negocio hasta su conclusión.*

Transferencia de custodia: Procedimiento habitual de ingreso de fondos en un archivo mediante traslado de las fracciones de series documentales, una vez que éstas han cumplido el plazo de permanencia fijado por las normas establecidas en la valoración para cada una de

las etapas del ciclo vital de los documentos. En la administración electrónica implica la remisión, según el modelo OAIS de paquetes de información (SIP) desde el organismo productor (o quien le haya sucedido en sus funciones) a una institución de archivo quien los conservará para su uso por la propia administración y los ciudadanos en forma de Paquetes de Información Archivística (AIP)

Trazabilidad: Propiedad o característica consistente en que las actuaciones de una entidad (persona o proceso) pueden ser imputadas exclusivamente a dicha entidad. En el ámbito de la gestión de documentos, proceso que facilita el seguimiento de la creación, incorporación, movimiento, uso y eventual modificación de los documentos dentro de un sistema de gestión de documentos.

Unidad organizacional: Parte identificada de una organización que despliega uno o más procesos que operan dentro de un conjunto coherente de objetivos de negocio y que forma la base para el alcance de una evaluación

Valoración: Fase del tratamiento archivístico que consiste en analizar y determinar los *valores primarios* y *secundarios* de las series documentales u otros *activos de información* (como es el caso de las bases de datos), fijando los plazos de transferencia, acceso, y conservación o eliminación total o parcial.

I. ***Valor primario:*** Aquél que va unido a la finalidad inmediata por la cual el documento se ha producido por la institución de carácter fiscal, judicial, jurídico-administrativo. etc.

   a) *Valor administrativo*: Aquél que posee un documento para la Administración de origen o aquella que le sucede como testimonio de sus procedimientos y actividades.

   b) *Valor Contable:* Aquel que tienen los documentos que pueden servir de explicación o justificación de operaciones destinadas al control presupuestario.

   c) *Valor Fiscal:* Aquel que tienen los documentos que pueden servir de testimonio del cumplimiento de obligaciones tributarias.

   d) *Valor Jurídico:* Aquel del que se derivan derechos u obligaciones legales regulados por el derecho común.

   e) *Valor Legal:* Aquél que pueden tener todos los documentos que sirvan de testimonio ante la ley.

II. ***Valor Secundario:*** Aquel que obedece a otras motivaciones que no son la propia finalidad del documento, tales como el valor histórico e informativo.

   a) *Valor Histórico:* Aquél que posee el documento como fuente primaria para la historia.

   b) *Valor Informativo:* Aquél que sirve de referencia para la elaboración o reconstrucción de cualquier actividad de la Administración y que también puede ser testimonio de la memoria colectiva.

10.2. Acrónimos

ACID: *Propiedad de las bases de datos, acrónimo formado por las iniciales de atomicidad (Atomicity), consistencia (Consistency), individualidad (Isolation) y durabilidad (Durability).*

AGE: Administración General del Estado.

AIP: Archival Information Package o Paquete de Información de Archivo

API: Application Programming Interface]

BSON: (*Binary JSON*) JSON Binario

CAP (Consistency, Availablilty, Partition Tolerance), teorema sobre la aplicación de las propiedades ACID en sistemas distribuidos.

CCN-CERT: Capacidad de Respuesta a Incidentes de Seguridad de la Información del Centro Criptológico Nacional, CCN

CCSDS: Consultative Committee for Space Data Systems; Comité Consultivo para Sistemas de Datos Espaciales

CERT[18]: (**C**omputer **E**mergency **R**esponse **T**eam**)** Equipo de Respuesta ante Emergencias Informáticas

CI: *Content Information; información de contenido,*

CSV: Denominación de extensión de fichero de texto donde los valores de las filas y columnas de una tabla, se almacenan usando caracteres de separación como comas, tabuladores o símbolos. Los Valores Separados por Comas (.CSV) y los Valores Separados por Tabulador (.TSV) son dos formatos de archivo usados habitualmente para exportar datos desde las bases de datos.

DAF: *Data Audit Framework; Marco de Auditoría de Datos*

DBMS (*database management system*): Sistema de gestión de bases de datos

DBPTK: Database Preservation Tool Kit

DI: *Descriptive Information; información descriptiva*

DIP: Dissemination Information Package: paquete de información de difusión

DRAMBORA: Digital Repository Audit Method Based on Risk Assessment; Metodología de Auditoría de Repositorios Digitales basada en el análisis de riesgos.

e-EMGDE: Esquema de Metadatos para la Gestión del Documento Electrónico.

ENI: Esquema Nacional de Interoperabilidad.

ENS: Esquema Nacional de Seguridad.

---

[18] *También se puede utilizar el término **CSIRT** (**C**omputer **S**ecurity **I**ncident **R**esponse **T**eam, Equipo de Respuesta ante Incidencias de Seguridad Informáticas) para referirse al mismo concepto*

ETL: *Extract Transform Load*

FI: *Fixity Information; información de integridad*

JSON: LOB: Large Object; Objeto de gran tamaño

LOPDGDD: Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y Garantía de los Derechos Digitales

ML: Machine learning

NERD: *Named Entity Recognition and Disambiguation*

NLP: natural language processing

NLU: *Natural Language Understanding*

NoSQL (Not only Structured Query Language):Tipo de bases de datos “no relacionales”

NTI: Norma Técnica de Interoperabilidad

OAIS (Open Archival Information System) es uno de los modelos de referencia impulsados por el Comité Consultivo para sistemas de datos del Espacio (CCSDS)

PDI: *Preservaion Description Information; información de descripción de la Conservación*

PGDe: Política de Gestión de Documentos Electrónicos

PI: *Packagin Information; Información de empaquetado*

PREMIS: *Preservation Metadata: Implementation Strategies; Grupo de trabajo para la implantación de estrategias para la conservación de Metadatos. Como resultado de su trabajo hay un Diccionario de Datos para los metadatos de conservación.*

REC: Registro Electrónico Común.

RGPD: Reglamento General de Protección de Datos; REGLAMENTO (UE) 2016/679 DEL PARLAMENTO EUROPEO Y DEL CONSEJO de 27 de abril del 2016

RI: *Reference Information; información de referencia*

SGBD: Sistema de Gestión de Base de Datos

SIARD: Software Independent Archiving of Relational Databases; es un formato de conservación de bases de datos, no vinculado a ningún proveedor de software

SICRES: Sistema de Información Común de Registros de Entrada y Salida

SIP: Submission Information Package: paquete de información de envío

SIR: Sistema de Interconexión de Registros.

SPARQL: (acrónimo recursivo ***S****PARQL* ***P****rotocol* ***a****nd* ***R****DF* ***Q****uery* ***L****anguage*.) Se traduce por un lenguaje estandarizado para la consulta de grafos RDF,

SQL: (*Structured Query Language*) lenguaje estructurado de consulta

TRAC: *Trusted Repository Archiving Checklist: Lista de comprobación para repositories de Archivo de Confianza.*

TTPR: *trusted third party repository; repositorio de tercero de confianza, RTC*

XML: eXtensible Markup Language.

XSD: XML Schema Definition.

YAML: Originariamente *Yet Another Markup Language* (otro lenguaje de marcado más). Tras su evolución se considera un acrónimo recursivo negando su naturaleza de lenguaje de marcado ***YAML Ain't Markup Language*** (‘YAML no es un lenguaje de marcado’)

11. REFERENCIAS

i. Ley 40/2015, de 1 de octubre, de Régimen Jurídico del Sector Público.

BOE-A-2015-10566

ii. Ley 39/2015, de 1 de octubre, del Procedimiento Administrativo Común de las Administraciones Públicas.

BOE-A-2015-10565

iii. REGLAMENTO (UE) 2016/679 DEL PARLAMENTO EUROPEO Y DEL CONSEJO de 27

BOE-A-2008-979

iv. Real Decreto 311/2022, de 3 de mayo, por el que se regula el Esquema Nacional de Seguridad.

BOE-A-2022-7191

v. Real Decreto 4/2010, de 8 de enero, por el que se regula el Esquema Nacional de Interoperabilidad en el ámbito de la Administración Electrónica.

BOE-A-2010-1331

vi. Real Decreto 203/2021, de 30 de marzo, por el que se aprueba el Reglamento de actuación y funcionamiento del sector público por medios electrónicos.

BOE-A-2021-5032

vii. Resolución de 28 de junio de 2012, de la Secretaría de Estado de Administraciones Públicas, por la que se aprueba la Norma Técnica de Interoperabilidad de Política de gestión de documentos electrónicos.

BOE-A-2012-10048

viii. Resolución de 19 de julio de 2011, de la Secretaría de Estado para la Función Pública, por la que se aprueba la Norma Técnica de Interoperabilidad de Expediente Electrónico.

BOE-A-2011-13170

ix. Resolución de 19 de julio de 2011, de la Secretaría de Estado para la Función Pública, por la que se aprueba la Norma Técnica de Interoperabilidad de Documento Electrónico.

BOE-A-2011-13169

x. Resolución de 19 de julio de 2011, de la Secretaría de Estado para la Función Pública, por la que se aprueba la Norma Técnica de Interoperabilidad de Digitalización de Documentos

BOE-A-2011-13168

xi. Resolución de 19 de julio de 2011, de la Secretaría de Estado para la Función Pública, por la que se aprueba la Norma Técnica de Interoperabilidad de Procedimientos de copiado auténtico y conversión entre documentos electrónicos.

BOE-A-2011-13172

xii. Resolución de 27 de octubre de 2016, de la Secretaría de Estado de Administraciones Públicas, por la que se aprueba la Norma Técnica de Interoperabilidad de Política de Firma y Sello Electrónicos y de Certificados de la Administración.

BOE-A-2016-10146

xiii. Decisión de Ejecución UE 2015/1506 de la Comisión, de 8 de septiembre de 2015, por la que se establecen las especificaciones relativas a los formatos de las firmas electrónicas avanzadas y los sellos avanzados que deben reconocer los organismos del sector público

DOUE-L-0-0037-00041

11.1. Estándares y buenas practicas

UNE-ISO/TR 15801:2019 IN Gestión de documentos. Información almacenada electrónicamente. Recomendaciones sobre confiabilidad y fiabilidad

UNE-ISO 16363:2017 Sistemas de transferencia de información y datos espaciales. Auditoría y certificación de repositorios digitales de confianza

UNE-ISO 17068 Repositorio de tercero de confianza para documentos electrónicos

UNE-ISO/TR 18128:2014 IN Información y documentación. Apreciación del riesgo en procesos y sistemas de gestión documental.

ISO 2709:2008 Information and documentation — Format for information exchange

ISO 8777:1993 Information and documentation — Commands for interactive text searching

ISO (2016) ISO/IEC 9075-1:2016 Information technology — Database languages — SQL — Part 1: Framework (SQL/Framework). Accesible en:

- https://web.archive.org/web/20210114164841
- https://www.iso.org/standard/63555.html

ISO/TR 10255:2009 Document management applications — Optical disk storage technology, management and standards

ISO/IEC 11179-1:2023 Information technology — Metadata registries (MDR) — Part 1: Framework

ISO/IEC 11179-3:2023 Information technology — Metadata registries (MDR) — Part 3: Metamodel for registry common facilities

ISO/IEC 11179-4:2004 Information technology — Metadata registries (MDR) — Part 4: Formulation of data definitions

ISO/IEC 11179-5:2015 Information technology — Metadata registries (MDR) — Part 5: Naming principles

ISO/IEC 11179-6:2023 Information technology — Metadata registries (MDR) — Part 6: Registration

ISO/IEC 11404:2007 Information technology — General-Purpose Datatypes (GPD)

ISO 12620-1:2022 Management of terminology resources — Data categories — Part 1: Specifications

ISO 12620-2:2022 Management of terminology resources — Data categories — Part 2: Repositories

ISO/TR 14105:2011 Document management — Change management for successful electronic document management system (EDMS) implementation

ISO 14289-1:2014 Document management applications — Electronic document file format enhancement for accessibility — Part 1: Use of ISO 32000-1 (PDF/UA-1)

ISO/DIS 14289-2 Document management applications — Electronic document file format enhancement for accessibility — Part 2: Use of ISO 32000-2 (PDF/UA-2)

ISO 14721:2012. Space data and information transfer systems – Open archival information system (OAIS) – Reference model. International Standards Organization. Ginebra: Suiza. ISO

ISO/PAS 15339-1:2015 Graphic technology — Printing from digital data across multiple technologies — Part 1: Principles

ISO/CD 15339-1.2 Graphic technology — Printing from digital data across multiple technologies — Part 1: Principles

ISO/PAS 15339-2:2015 Graphic technology — Printing from digital data across multiple technologies — Part 2: Characterized reference printing conditions, CRPC1-CRPC7

ISO 15801 Electronic imaging – Information stored electronically – Recommendations for trustworthiness and reliability.

ISO 15836-1:2017 Information and documentation — The Dublin Core metadata element set — Part 1: Core elements.

ISO 15836-2:2019 Information and documentation — The Dublin Core metadata element set — Part 2: DCMI Properties and classes.

UNE-ISO 16175-1:2012. Información y documentación. Principios y requisitos funcionales para documentos en entornos de oficina electrónica. Parte 1: Generalidades y declaración de principios. Madrid: AENOR, 20 p.

UNE-ISO 16175-2:2012. Información y documentación. Principios y requisitos funcionales para documentos en entornos de oficina electrónica. Parte 2: Directrices y requisitos funcionales para sistemas que gestionan documentos electrónicos. Madrid: AENOR, 54 p.

UNE-ISO 16175-3:2012. Información y documentación. Principios y requisitos funcionales para documentos en entornos de oficina electrónica. Parte 3: Directrices y requisitos funcionales para documentos en los sistemas de la organización. Madrid: AENOR, 68 p

ISO 16642:2017 Computer applications in terminology — Terminological markup framework

ISO 17933:2000 GEDI — Generic Electronic Document Interchange

ISO/TS 18759:2022 Document management — Trustworthy storage system (TSS) — Functional and technical requirements

ISO 20614:2017 Information and documentation — Data exchange protocol for interoperability and preservation

ISO 20775:2009 Information and documentation — Schema for holdings information

ISO 21127:2014 Information and documentation — A reference ontology for the interchange of cultural heritage information

ISO/IEC 21320-1:2015 Information technology — Document Container File — Part 1: Core

ISO/TR 22299:2018 Document management — Digital file format recommendations for long-term storage

ISO 23504-1:2020 Document management applications — Raster image transport and storage — Part 1: Use of ISO 32000 (PDF/R-1)

ISO 24517-1:2008 Document management — Engineering document format using PDF — Part 1: Use of PDF 1.6 (PDF/E-1)

UNE-ISO 23081-1:2018 Información y documentación. Procesos de gestión de documentos. Metadatos para la gestión de documentos. Parte 1: Principios.

UNE-ISO 23081-2:2021 Información y documentación. Procesos de gestión de documentos. Metadatos para la gestión de documentos. Parte 2: Elementos conceptuales y de implementación.

UNE-ISO/TR 23081-3:2012 IN Información y documentación. Metadatos para la gestión de documentos. Parte 3: Método de auto-evaluación.

ISO/IEC 27557:2022 Information security, cybersecurity and privacy protection — Application of ISO 31000:2018 for organizational privacy risk management

UNE-ISO 31000:2018 Gestión del riesgo. Directrices

BS 31100:2021 Risk management. Code of practice and guidance for the implementation of BS ISO 31000:2018

BS 31100:2021 – TC Tracked Changes. Risk management. Code of practice and guidance for the implementation of BS ISO 31000:2018

BS 31100:2021 ExComm Expert Commentary for BS 31100:2021. Risk management. Code of practice and guidance for the implementation of BS ISO 31000:2018

Australasian Digital Recordkeeping Initiative (ADRI) https://www.caara.org.au/index.php/working-groups/adri/#:~:text=ADRI%20was%20formed%20to%20articulate,made%20accessible%20for%20the%20future.

Australian Government Recordkeeping Metadata Standard Implementation Guidelines

National Archives of Australia.V 2.0 June 2011 https://www.naa.gov.au/sites/default/files/2019-09/agrkms-implementation-guidelines_tcm16-93992_0.doc

Australian Government Recordkeeping Metadata Standard (AGRkMS) National Archives of Australia June 2015 V 2.2 https://www.naa.gov.au/sites/default/files/2019-09/AGRkMS-Version-2.2-June-2015_tcm16-93990_1.pdf

DCC Curation Lifecycle Model: from the Digital Curation Centre

Digital Curation Manual: from the Digital Curation Centre

Digital Preservation Handbook: from the Digital Preservation Coalition (DPC)

Directivas sobre Documentos electrónicos: Transition of Electronic Records (M-19-21)

KEEP Solutions (2020) DBPTK (Database Preservation Toolkit). Accesible en:

- https://web.archive.org/web/20210112095121
- https://database-preservation.com/

Levels of Digital Preservation: from the National Digital Stewardship Alliance (NDSA)

Library of Congress [LC] (2020-2021) Recommended formats statement: Datasets. Accesible en:

- https://web.archive.org/web/20201116041423
- http://www.loc.gov/preservation/resources/rfs/dat a.html

Library of Congress (2017). SQLite, Version 3. Accesible en:

- https://web.archive.org/web/20201117025418
- https://www.loc.gov/preservation/digital/formats/f dd/fdd000461.shtml

Library of Congress (2015) SIARD (Software Independent Archiving of Relational Databases) Version 1.0. Accesible en:

- https://web.archive.org/web/20201101014106
- https://www.loc.gov/preservation/digital/formats/f dd/fdd000426.shtml

National Archives of Australia (2020): *Digital Preservation Policy https://www.naa.gov.au/about-us/our-organisation/accountability-and-reporting/archival-policy-and-planning/digital-preservation-policy*

OAIS Reference Model: introduction to the Open Archival Information System Reference Model

RDB SIARD (2019) Preserving databases using SIARD: Experiences with workflows and documentation practices: CEF eArchiving Building Block, E-ARK3 [CEF]. Accesible en:

- https://web.archive.org/web/20201230133036
- https://dilcis.eu/images/2020review/9_Draft_SIARD_Case_Study_1.pdf

Software Preservation Network (2020) Software Preservation Network. Accesible en:

- https://web.archive.org/web/20210107085536
- https://www.softwarepreservationnetwork.org/

Software Sustainability Institute (2020) Digital preservation and curation - the danger of overlooking software. Accesible en:

- https://web.archive.org/web/20191128143837
- https://www.software.ac.uk/resources/guides/digital-preservation-and-curation-danger-overlooking-software

Swiss Federal Archives (2013-03-21) eCH-0165 SIARD-Formatspezifikation. Accesible en

- https://web.archive.org/web/20201028110949
- https://www.ech.ch/de/dokument/2760f452-6e56- 48ef-bb0d-20f68638a825

Swiss Federal Archives (2020) “SIARD Suite” -- Tools -- Archiving. Accesible en:

- https://web.archive.org/web/20201030171732
- https://www.bar.admin.ch/bar/en/home/archiving/tools/siard-suite.html

The National Archives (n.d.) Digital Object Record Identification (DROID). Accesible en:

- https://web.archive.org/web/20201015033155
- https://github.com/digital-preservation/droid

The National Archives (2020) The Technical Registry: PRONOM. Accesible en:

- https://web.archive.org/web/20201111032324
- http://www.nationalarchives.gov.uk/PRONOM/Default.aspx

The National Archives (2010) PRONOM: SIARD (Software-Independent Archiving of Relational Databases) 1.0. Accesible en:

- https://web.archive.org/web/20201101173235
- https://www.nationalarchives.gov.uk/PRONOM/fmt/161

Consejo Superior de Informática y para el Impulso de la Administración Electrónica (2004): *Criterios de seguridad, normalización y conservación de las aplicaciones utilizadas en el ejercicio de potestades.* Ministerio de Administraciones Públicas. NIPO 326-04-044-9. Accesible en: https://administracionelectronica.gob.es/pae_Home/pae_Biblioteca/pae_PublicacionesPropias/Monografias-administracion-electronica/Criterios_seguridad_normalizacion_conservacion.html?comentarioContenido=0

11.2. Bibliografía

- Archaeology Data Service (2009) Databases and Spreadsheets: A Guide to Good Practice. Accesible en:
  - https://web.archive.org/web/20201215095638
  - https://guides.archaeologydataservice.ac.uk/g2gp/DbSht_Toc
- *BCS Academy Glossary Working Party* (2013) BCS Glossary of Computing and ICT 13th edition. Accesible en:
  - https://learning.oreilly.com/library/view/bcs-glossaryof/9781780171500/11_GlossaryofICT_partA9.xhtml [accessed 24 March 2021]
- Claris (2021) Use *Claris* FileMaker to Build Business Applications — Claris. Accesible en: https://web.archive.org/web/20210101004554/https://www.claris.com/filemaker/
- COPTR (2021) Database. Accesible en: https://web.archive.org/web/20210706064251/https://coptr.digipres.org/index.php/Database
- Digital Preservation Coalition (2015): Digital Preservation Handbook, 2nd Edition https://www.dpconline.org/handbook.
- DLM Forum (1997): Guidelines on best practices for using electronic information. How to deal with machine-readable data and electronic documents. Luxembourg. Office for Official Publications of the European Communities.
- E-ARK (2017a) Welcome to the E-ARK Project. Accesible en: https://web.archive.org/web/20201218052540/http://eark-project.com/

- E-ARK (2017b) SIARD 2.0. Accesible en: https://web.archive.org/web/20200925185826/https://earkproject.com/resources/specificationdocs/32-specification-for-siard-formatv20/STAN_e_FINAL_2015-07-04_eCH-0165_V2%200_SIARD-Format.pdf
- Freitas R. and Ramalho, J. (2009) Relational Databases Digital Preservation. Accesible en: http://xata.fe.up.pt/2009/papers/79.pdf
- Haerder, T. and Reuter, A. (1983). Principles of transaction-oriented database recovery. Accesible en: https://doi.org/10.1145/289.291 . DOI: 10.1145/289.291
- Harrison, Guy (2015): Next Generation Databases NoSQL, NewSQL and Big Data, Guy Harrison, Apress,
- Heslop, Helen & Davis, Simon & Wilson, Andrew & National Archives of Australia (2002). National Archives green paper: an approach to the preservation of digital records. NAA, [Canberra].
- Keefer Riva, A. (2014). Preservación de recursos de información digital. Barcelona: UOC.
- Lehane, R (2020) Siegfried. Accesible en:
    - https://web.archive.org/web/20201028192837
    - https://github.com/richardlehane/siegfried
- Microsoft (2016) Database Snapshots (SQL Server). Accesible en:
    - http://web.archive.org/web/20210506145800
    - https://docs.microsoft.com/en-us/sql/relationaldatabases/databases/database-snapshots-sql-server?view=sql-server-ver15
- Microsoft (2021a) Database Software and Applications | Microsoft Access. Accesible en:
    - https://web.archive.org/web/20201223120917
    - https://www.microsoft.com/en-us/microsoft365/access
- Microsoft (2021b) Introduction to importing, linking, and exporting data in Access. Accesible en:
    - https://web.archive.org/web/20201109011756
    - https://support.microsoft.com/enus/office/introduction-to-importing-linking-and-exporting-data-in-access-08422593-42dd-4e73-bdf1-4c21fc3aa1b0?ui=en-us&rs=en-us&ad=us
- MILARUELO, ALEJANDRO (2020) “Conservación a largo plazo de bases de datos” en La gestión del documento electrónico, Wolters & Kluwer.
- Open Preservation Foundation (2020) Format Identification for Digital Objects (FIDO). Accesible en:
    - https://web.archive.org/web/20200916134739
    - https://github.com/openpreserve/fido

- Oracle (2020) Oracle Database. Accesible en:
    - https://web.archive.org/web/20210105064333
    - https://www.oracle.com/database/
- Proceedings of Workshop: Databases for 2080 – Preserving database content for the long term (2021). **Landesarchiv Baden-Württemberg** accessible (20221104) en https://nbn-resolving.org/urn:nbn:de:101:1-2022071903
- Rothenberg, Jeff (1995) “Ensuring the Longevity of Digital Documents” en Scientific American (Vol. 272, Number 1, pp. 42-7; January 1995).
- Rothenberg, Jeff (1998) *Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation* . Reporter for the Council on Library and Information Resources (CLIR), January 1998. ISBN 1-887334-63-7. Versión PDF accesible en http://www.clir.org/pubs/reports/rothenberg/pub77.pdf
- Rothenberg, Jeff (1999) Ensuring the Longevity of Digital Information (Wahington, DC: Council on Library and Information Resouces) accessible en https://www.clir.org/wp-content/uploads/sites/6/ensuring.pdf .
- Rothenberg, Jeff (2000): Preserving Authentic Digital Information . En report Authenticity in a Digital Environment, CLIR (ISBN 1-887334-77-7), pp. 51-68. Accesible en http://www.clir.org/pubs/reports/pub92/pub92.pdf
- Thomson, Sara Day (2016) Preserving Transactional Data. Accesible en:
    - http://dx.doi.org/10.7207/twr16-02. DOI: 10.7207/twr16-02.
- Wikipedia (2021a) Databases. Accesible en:
    - https://web.archive.org/web/20210108180820
    - https://en.wikipedia.org/wiki/Database
- Wikipedia (2021b) Oracle Database. Accesible en:
    - https://web.archive.org/web/20201218205237
    - https://en.wikipedia.org/wiki/Oracle_Database

Equipo de trabajo de Gestión y conservación del dato Grupo de trabajo de Política de gestión de documentos electrónicos (Comisión Ministerial de Administración Digital)	
Presidente	
Gerardo Bustos Pretel	Secretaría General Técnica (MINHAC)
Equipo de redacción	
José Luis García Martínez	Secretaría General Técnica (MINHAC)
Anselmo Fajardo Alonso	Secretaría General Técnica (MINHAC)
Luis José Megino Collado	Secretaría General Técnica (MINHAC)
Equipo de trabajo de gestión y conservación del dato	
Carmen García Roger	Inspección General (MINHAC)
Andoni Pérez de Lema Sáenz de Viguera	IGAE (MINHAC)
Carlos Alonso Peña	Oficina del dato (MTDFP)
Alma González Correa	Oficina del Dato (MTDFP)
Antonio Sanz Pulido	Inspección General (MINHAC)
Alejandro Millaruelo Gómez	IGAE (MINHAC)
Ainara Toca Muguira	Secretaría General de Administración Digital (MTDFP)
Equipo de revisión jurídica	
María Bueyo Díez-Jalón	Abogacía del Estado (MINHAC)
Alejandra González Madrid	Abogacía del Estado (MINHAC)
Participación en el grupo de trabajo	
Miguel Ángel Amutio Gómez	Secretaría General de Administración Digital (MTDFP)
Josefina Otheo de Tejada Barasoain	AEAT
Julián Antonio Prior Cabanillas	Dirección General de Gobernanza Pública (MINHAC)
Ana María Porras del Río	Secretaría de Estado de Función Pública (MINHAC)
Beatriz Bernáldez Méndez	Parque Móvil del Estado (MINHAC)
Javier Guerra Casanova	SGTIC (MINHAC)
Elena Galdón Torrezano	MPT
Manuel Gómez Vaz	INAP (MTDFP)
María Berenguer Carretero	Secretaría de Estado de Función Pública (MINHAC)
Manuel Mesa de Santiago	Oficina del dato (MTDFP)
Cándido Peña	AEAT
David Gómez Castellanos	Secretaría General de Administración Digital (MTDFP)
Carlos López Orcajo	DG Tributos (MINHAC)
Aurelio Lozano Serra	MINHAC
Juan Pablo García	AEAT
Elsa Hernández Fernández	Secretaría de Estado de Función Pública (MINHAC)
Carmen Santamaría Barceló	Secretaría General Técnica (MINHAC)
Miguel Ángel Blanco Arribas	Secretaría General de Financiación Autonómica y Local (MINHAC)
Jesús Pardo Ballenato	Fábrica Nacional de Moneda y Timbre
Caridad Gascón Soriano	Secretaría de Estado de Presupuestos y Gastos (MINHAC)
Robert de Jorge Domingo	Secretaría de Estado de Hacienda (MINHAC)
José Lino Monteagudo Pereira	IGAE (MINHAC)
Francisco Javier Garrido Gómez	Fábrica Nacional de Moneda y Timbre
Sandra Matesanz Guerra	AEAT
Elena Galdón Torrezano	DTIC (MPT)
María Teresa Villaizán Montoya	Delegación de Economía y Hacienda de Castilla y León (MINHAC)
Sonia María Cascales Sedano	Dirección General del Catastro (MINHAC)
Enrique Ladrón de Guevara	Secretaría de Estado de Función Pública (MINHAC)
Marta Hidalgo Arroyo	Dirección General del Catastro (MINHAC)
Víctor Balbás Valderrábano	Secretaría General de Administración Digital (MTDFP)
Rafael Rojas Corral	Instituto de Estudios Fiscales (MINHAC)
Víctor Casado Izquierdo	TEAC (MINHAC)
David Lleras Iglesias	Subsecretaría (MPT)
María de la Plaza Villaroya	SGTIC (MINHAC)
María Isabel de la Rosa	Secretaría de Estado de Función Pública (MTDFP)