

## ESTRATEGIAS PARA LA PRESERVACIÓN DIGITAL A LARGO PLAZO DE LOS DOCUMENTOS DE ARCHIVO ELECTRÓNICOS

### Objetivo

1. Las presentes estrategias tienen por objeto prevenir el deterioro de los documentos de archivo electrónico y garantizar la permanencia de la información que se localiza en medios digitales, sin generar alteraciones al soporte y/o información, aplicando los criterios de autenticidad, integridad, confidencialidad, inalterabilidad, fiabilidad, interpretación, comprensión y disponibilidad conforme a sus plazos de conservación señalados en el Catálogo de Disposición Documental del Instituto Nacional Electoral.

### Ámbito de aplicación

2. Son de carácter general y están dirigidas a todo el personal del Instituto, así como a las personas Prestadoras de Servicios Eventuales y Permanentes.

### Glosario

3. Para los efectos de este documento se entenderá por:

#### A. Siglas y Abreviaturas

- I. **Instituto:** Instituto Nacional Electoral;
- II. **PNG/PNGX:** Portable Network Graphic (Gráfico de red portátil);
- III. **SAI:** Sistema de Archivos Institucional;
- IV. **UTSI:** Unidad Técnica de Servicios de Informática;
- V. **UTTyPDP:** Unidad Técnica de Transparencia y Protección de Datos Personales

#### B. Definiciones

- I. **Autenticidad:** La característica del documento cuando puede probar que es lo que afirma ser, que ha sido creado o enviado por la persona que se afirma que lo ha creado o enviado y que ha sido creado o enviado en el momento que se declara;
- II. **Conservación:** El conjunto de medidas preventivas o correctivas adoptadas para garantizar la integridad física de los documentos de archivo, sin alterar su contenido;
- III. **Disponibilidad:** Característica del documento electrónico y sus metadatos que lo hace accesible y utilizable bajo demanda por las personas autorizadas;
- IV. **DOC, DOCX:** Extensión de nombre de archivo utilizada para documentos de procesamiento de texto almacenados en el formato de archivo binario de Microsoft Word;
- V. **Documento electrónico:** Es aquel documento que es generado, consultado, modificado o procesado por medios electrónicos;

- VI. Hash:** es un valor o código de representación único que permite identificar a cada documento electrónico o digitalizado. Dicho hash o código es obtenido a partir de algún algoritmo criptográfico denominado comúnmente funciones hash.
- VII. Integridad:** El carácter de un documento de archivo que es completo y veraz y refleja con exactitud la información contenida;
- VIII. JPG/JPGX:** Formato de archivo estándar utilizado para la comprensión de imágenes, tanto en color como en escala de grises de alta calidad;
- IX. Metadatos:** Conjunto de datos estructurados que describen las propiedades de un documento electrónico y son de carácter instrumental e independiente de su contenido, destinados al conocimiento inmediato y automatizable de algunas de sus características, con la finalidad de garantizar la disponibilidad, el acceso, la conservación y la interoperabilidad; sirven para documentar, estructurar, interpretar, identificar, procesar, recuperar, preservar, usar, administrar, vincular y presentar un conjunto de datos específicos;
- X. Migración:** Serie de tareas organizadas para la transferencia periódica de información digital desde un sistema a otro, desde un formato a otro, de un sistema operativo a otro, o desde un lenguaje de programación a otro;
- XI. Obsolescencia tecnológica:** Es la dificultad de acceder a la información contenida en un documento electrónico debido a que el formato utilizado ha caído en desuso y no es compatible con el *software* utilizado en la tecnología actual, resultando inadecuado para las necesidades actuales de almacenamiento y/o conservación de documentos electrónicos de archivo;
- XII. PDF:** Formato estándar de documento portátil que permite presentar e intercambiar documentos de manera fiable, independientemente del *software*, el *hardware* o el sistema operativo;
- XIII. Preservación a largo plazo:** Conjunto de acciones y estándares aplicados a los documentos durante su gestión para garantizar su preservación en el tiempo, independientemente de su medio y forma de registro o almacenamiento. La preservación a largo plazo aplica al documento de archivo electrónico con su medio correspondiente en cualquier etapa de su ciclo vital;
- XIV. Preservación digital:** El proceso específico para mantener los materiales digitales durante las diferentes generaciones de la tecnología, a través del tiempo, con independencia de los soportes en los que se almacenan;
- XV. Software:** Conjunto de programas de cómputo, procedimientos, reglas, documentación y datos asociados, que forman parte de las operaciones de un sistema informático;
- XVI. TIFF:** Formato estándar de archivo de imagen que se utiliza para almacenar imágenes digitales de alta calidad, permitiendo su compatibilidad con diferentes aplicaciones y/o sistemas operativos;
- XVII. XLS/XLSX:** Formato de archivo informático utilizado para almacenar hojas de cálculo electrónicas;

### Principios de preservación digital a largo plazo

4. No es posible preservar un documento electrónico de archivo, si no se contextualiza su producción a través de la aplicación de principios documentales como son:
- a) **Principio de Procedencia:** el cual establece que se deberá conservar el origen de cada fondo documental producido por los sujetos obligados a la normatividad en materia de archivos, para distinguirlo de otros fondos semejantes y respetar el orden interno de las series documentales en el desarrollo de su actividad institucional; y
  - b) **Principio de Orden Original:** el cual establece que la disposición de los documentos debe respetar la secuencia de los trámites que los produjo. Es prioritario para la ordenación de fondos, series y unidades documentales.

No se puede preservar un expediente, si no se han resguardado los tipos documentales electrónicos que lo integran, en el marco de una relación con otros documentos que, en conjunto, lo conforman.

5. Las principales estrategias de preservación que se deberán aplicar al archivo digital son:
- a) **Migración:** Cambio a nuevos formatos/plataformas.
  - b) **Replicado:** Copias de la información digital establecidas según las políticas y recomendaciones de conservación de la información.
  - c) **Normalización de formatos:** Cuando se determina el uso de formatos normalizados y/o estandarizados, desde la creación de los documentos, se minimiza el uso y aplicación de técnicas de preservación como la migración.

### Estrategias

#### I. Identificar los documentos de archivo electrónicos susceptibles de preservación digital a largo plazo

La estrategia está dirigida a las áreas responsables de manejar o custodiar información, a efecto de que puedan identificar los tipos documentales a preservar a largo plazo, ya que no todos los documentos producidos y recibidos en el Instituto Nacional Electoral, son susceptibles de ser preservados.

Dicha estrategia consistirá en la identificación de los documentos de archivo electrónicos a preservar por más 10 años, según lo establece el Catálogo de Disposición Documental.

**Actividades por realizar:**

1. Identificar los documentos de archivo electrónicos que, conforme al Catálogo de Disposición Documental aplicable, tienen un plazo de conservación igual o superior a 10 años<sup>1</sup> o de conservación permanente.
2. Identificar la información a preservar de los activos de información. En un expediente que se conservará de manera permanente por poseer valores secundarios, puede ser que no se conserve todo el expediente, sino un muestreo de los documentos más relevantes que lo conforman, selección que se realiza en el proceso de disposición final por el Archivo Institucional.
3. Identificar los medios de almacenamiento en los que se encuentran los documentos de archivo electrónicos.
4. Identificar las aplicaciones o software por las cuales se producen documentos en medio electrónico.
5. Realizar el inventario de documentos electrónicos de archivos, en el cual se determinará formato del documento, número de páginas, volumen de cada archivo, peso, versión, nombre del expediente, metadatos, entre otros.

Para la referencia de metadatos se podrá usar el modelo OAIS (Open Archival Information System) que establece metadatos: descriptivos, administrativos, técnicos, estructurales, de preservación, etc.

Algunos ejemplos de metadatos son:

| No. | Metadato                                 | Descripción  | Tipo      |
|-----|--|--|-----------|
| 1   | Título del documento.                    | Es el nombre que se le asigna al documento. Se sugiere que el número máximo de caracteres sea de 30.         | Editable. |
| 2   | Descripción del contenido del documento. | Descripción detallada del contenido del documento. Se sugiere que el número máximo de caracteres sea de 250. | Editable. |

<sup>1</sup> Plazo que se deberá computar una vez concluido el asunto del expediente que se trate.

| <b>No.</b> | <b>Metadato</b>                           | <b>Descripción</b>  | <b>Tipo</b>  |
|------------|---|---|--------------|
| 3          | Tipo de archivo.                          | El tipo de archivo electrónico (por ejemplo, PDF o PDFX, DOC o PDFX, o bien XLS o XLSX).  | Editable.    |
| 4          | Nombre de la persona autora.              | El nombre de la persona autora del documento original que ha sido digitalizado.<br><br>En caso de no conocerse el nombre de la persona autora, se colocará "Desconocido".   | Editable.    |
| 5          | Área de adscripción de la persona autora. | Se refiere a la Unidad Responsable a la cual pertenece el autor del documento original que ha sido digitalizado.<br><br>En caso de que la persona autora no pertenezca al INE, colocar "Externo".<br><br>En caso de no conocerse el nombre de la persona autora colocar "No aplica" "Se desconoce" según sea el caso. | Editable.    |
| 6          | Palabras clave.                           | Serie de etiquetas que describen el contenido del archivo.  | Editable.    |
| 7          | Nivel de acceso.                          | Definir si el acceso es público o privado.  | Editable.    |
| 8          | En su caso, fecha de digitalización.      | Es la fecha en la que se llevó a cabo la digitalización del documento.  | No editable. |
| 9          | Tamaño del archivo.                       | El tamaño del archivo electrónico.  | No editable. |
| 10         | Formato.                                  | Es el formato del documento.  | No editable. |
| 11         | Tipo de compresión utilizado.             | Se refiere al tipo de compresión que se utilizó para almacenar la información del documento.  | No editable. |

Se recomienda revisar y validar los metadatos de forma regular para garantizar su calidad y precisión.

6. La documentación de archivo electrónico que, conforme al Catálogo de Disposición Documental aplicable, tengan como destino final su conservación permanente, deberán encontrarse registrados y descritos de forma individual en el módulo e-archivo del Sistema de Archivos Institucional.

## II. Establecer los formatos de preservación digital a largo plazo

La estrategia está orientada a establecer formatos de creación de documentos o conversión, de ser posible, de los ya creados. Los documentos de archivo electrónicos del Instituto Nacional Electoral serán soportados en formatos de preservación a largo plazo y de uso común recomendados por diversas normas internacionales en preservación digital a largo plazo, los cuales se listan a continuación:

| <b>CONTENIDO EN TEXO</b>   |                |   |                  |
|----------------------------|----------------|---|------------------|
| <b>ESTÁNDAR</b>            | <b>FORMATO</b> | <b>CARACTERÍSTICAS</b>  | <b>EXTENSIÓN</b> |
| ISO 19005                  | PDF/A          | Formato de archivo de documentos electrónicos para la Preservación a largo plazo.   | .pdf             |
| W3C HTML Estandar abierto  | XML            | Es un estándar abierto, flexible y ampliamente utilizado para almacenar, publicar e intercambiar cualquier tipo de información. | .xml             |
| <b>CONTENIDO EN IMAGEN</b> |                |   |                  |
| <b>ESTÁNDAR</b>            | <b>FORMATO</b> | <b>CARACTERÍSTICAS</b>  | <b>EXTENSIÓN</b> |
| ISO/IEC 15444              | JPEG2000       | JPEG2000 (sin pérdida) permite reducir el peso de los archivos a la mitad en comparación con las imágenes no comprimidas.       | .jpg2 - jp2      |
| ISO 12639                  | TIFF           | TIFF (sin compresión) Archivos más grandes que un formato comprimido  | .tiff            |
| <b>CONTENIDO EN AUDIO</b>  |                |   |                  |
| <b>ESTÁNDAR</b>            | <b>FORMATO</b> | <b>CARACTERÍSTICAS</b>  | <b>EXTENSIÓN</b> |
| EBU - TECH 3285            | BWF            | Formato de archivo que toma la estructura de archivos WAVE existente y añade metadatos adicionales                              | .bwf             |
| <b>CONTENIDO EN VIDEO</b>  |                |   |                  |
| <b>ESTÁNDAR</b>            | <b>FORMATO</b> | <b>CARACTERÍSTICAS</b>  | <b>EXTENSIÓN</b> |

|             |                         |   |             |
|-------------|-------------------------|---|-------------|
| ISO 15444-4 | JPEG<br>2000-<br>Motion | Formato para la Preservación sin pérdida de vídeo en formato digital y migración de las grabaciones de vídeo analógicas obsoletos en archivos digitales | .mj2 -.mjp2 |
|-------------|-------------------------|---|-------------|

**CONTENIDO DE COMPRESIÓN**

| ESTÁNDAR                  | FORMATO | CARACTERÍSTICAS                | EXTENSIÓN |
|---------------------------|---------|--------------------------------|-----------|
| RFC 1952 Estándar Abierto | GZIP    | Formato de compresión de datos | .gz       |

**CONTENIDO EN BASE DE DATOS**

| ESTÁNDAR | FORMATO | CARACTERÍSTICAS  | EXTENSIÓN |
|----------|---------|--|-----------|
|          | SIARD   | Formato para el archivo de bases de datos relacionales en una forma independiente del proveedor, delimitado archivos planos (texto sin formato ) con DDL | .siard    |

**CONTENIDO PAGINAS WEB**

| ESTÁNDAR  | FORMATO        | CARACTERÍSTICAS   | EXTENSIÓN |
|-----------|----------------|---|-----------|
| ISO 28500 | Web<br>ARChive | Utilizado para almacenar "Web crawls" como secuencias de bloques de contenido recolectados de la World Wide Web | .warc     |

**CONTENIDO CORREO ELECTRÓNICO**

| ESTÁNDAR | FORMATO | CARACTERÍSTICAS  | EXTENSIÓN       |
|----------|---------|--|-----------------|
| RFC 822  | EML     | Diseñado para almacenar mensajes de correo electrónico en forma de un archivo de texto sin formato | .eml            |
|          | MBOX    | Formato utilizado para almacenar conjuntos de correos electrónicos.                                | .mbox -<br>.mbx |

**Actividades por realizar:**

1. Identificar los formatos de producción de información.
2. Elaborar matriz que coteje formatos actuales y los formatos propuestos para preservación.
3. Establecer en los sistemas existentes, la recomendación sobre el uso de los formatos de preservación a largo plazo, tanto para los ya creados, como para los que se generen en adelante a la aprobación de las presentes estrategias.

### **III. Normalizar la producción documental con formatos de archivos de preservación a largo plazo**

La estrategia está dirigida a las áreas responsables de manejar o custodiar información, orientándolas a efecto de regular, así como normalizar la producción y uso de los de los documentos de archivo electrónicos en formatos con características y atributos de estándares de preservación a largo plazo.

Para lo cual se deberá controlar las especificaciones y características de los formatos de los documentos de archivo electrónicos, mismas que deberán quedar debidamente registradas en las Fichas Técnicas de Valoración Documental y, en consecuencia, en las series documentales del Catálogo de Disposición Documental vigentes del Instituto Nacional Electoral.

#### **Actividad por realizar:**

1. Elaborar y aplicar el procedimiento para la normalización de la producción documental electrónica con formatos de preservación digital a largo plazo.

### **IV. Migrar a formatos de preservación a largo plazo**

La estrategia está dirigida a las áreas responsables de manejar o custodiar información y consiste en realizar el cambio a nuevos formatos que garanticen la preservación de la documentación de archivo electrónica a largo plazo de manera estandarizada; tales como PDF/A, XML, TIFF entre otros

#### **Actividades por realizar:**

1. De los documentos susceptibles de preservación a largo plazo identificar el formato de origen.
2. Identificar el formato apropiado para efectuar conversión.
3. Verificar la autenticidad de los documentos de archivo electrónicos, en su caso, mediante la revisión de los certificados de firma electrónica.
4. Realizar la conversión de los documentos de archivo electrónicos susceptibles de preservación que se encuentren en formatos no apropiados a su nuevo formato.
5. Realizar la aplicación de la función hash al documento migrado.



## **V. Renovación de dispositivos y medios**

La presente estrategia está dirigida a las áreas responsables de manejar o custodiar información, con apoyo del área de tecnologías de información que se trate, y consiste en la renovación periódica de los medios de almacenamiento con el fin de mitigar amenazas derivadas de los altos niveles de obsolescencia tecnológica que se presenta en los medios de almacenamiento de los documentos, sumado a la limitada durabilidad que tienen, la vulnerabilidad al deterioro y a pérdida por intrusiones humanas, fallas catastróficas o desastres naturales.

### **Actividades por realizar:**

1. Identificar los dispositivos y medios de almacenamiento de documentos de archivo electrónicos.
2. Identificar las edades de los medios de almacenamiento.
3. Elaborar y aplicar el programa de monitoreo y renovación de medios de almacenamiento.
4. Realizar copias de seguridad en distintas ubicaciones de manera periódica.
5. Sistematizar las copias de seguridad señaladas en el numeral anterior, de preferencia con un software que realice copias de manera automática y calendarizada, en horarios en los que no se encuentre en uso la información que sea susceptible de respaldarse, resguardándola en ubicaciones seguras para garantizar que se pueda acceder a ella.

Algunos de los tipos de copias de seguridad que se pueden utilizar en función de la necesidad, son los siguientes:

- a) **Copia de seguridad completa:** se recomienda elegir si se necesita un respaldo completo de toda la información, siendo la opción ideal si se quiere tener todo completamente protegido. Este tipo de copia requiere más tiempo y más espacio para llevarse a cabo.
- b) **Copia de seguridad diferencial:** solo incluye los archivos que se han cambiado desde la última vez que se realizó una copia previa, de forma que se incluirán los archivos nuevos. Opción ideal si ya se cuenta con una copia y solo se requiere actualizarla con nuevos datos o archivos modificados.
- c) **Copia de seguridad incremental:** si solo se requiere realizar una copia de los archivos que han sido modificados desde la última copia realizada, esta es la opción para elegir, siendo la más rápida.

6. Actualizar periódicamente los sistemas operativos y aplicaciones, para mitigar riesgos en materia de seguridad que van surgiendo.

## **VI. Implementar mecanismos de autenticidad e integridad de los documentos de archivo electrónicos**

Para garantizar que los documentos de archivo electrónicos se preserven sin alteraciones o manipulaciones indebidas en su contenido, contexto y estructura, las áreas responsables de generar documentación deben utilizar técnicas como: estampado cronológico, firmas electrónicas, certificados digitales, código seguro de verificación, marcas de agua digital.

### **Actividades por realizar:**

1. Validar en el sistema de gestión documental que los metadatos existen.
2. Implementar el uso de la Firma Electrónica Avanzada Institucional para preservar la seguridad de los documentos de archivo electrónico, ya que la misma garantiza la integridad y autenticidad de un documento que fue firmado electrónicamente.
3. Uso de códigos de integridad utilizando algoritmos robustos como mecanismo de comprobación de que los documentos de archivo electrónico no han sido modificados a través del tiempo; para lo cual se puede hacer uso del estándar SHA256.<sup>2</sup>

## **Estrategias técnicas de preservación digital**

### **I. Suficiente capacidad de almacenamiento**

Aunque esta capacidad puede ser añadida en cualquier momento es importante tener en cuenta que los sistemas informáticos (como el Sistema de Archivos Institucional) que resguardan los documentos de archivo electrónicos, deben tener espacio suficiente para almacenar los datos previstos durante todo el ciclo de vida de estos.

### **II. Arqueología digital**

En caso de daño o pérdida, aplicar un conjunto de métodos que permitan en la medida de lo posible recuperar la información a partir de fuentes dañadas, fragmentadas o arcaicas.

---

<sup>2</sup> Es un ejemplo de función HASH.

### **III. Continuidad y disponibilidad**

Implementar la infraestructura y/o servicios que permitan asegurar la operación de los sistemas y medios de almacenamiento a efecto de proveer el acceso a la información de forma continua.

#### **Beneficios de la preservación digital**

- a)** Contribuye al medio ambiente y en la disminución de la contaminación ambiental generada por el consumo de papel.
- b)** Adoptar estándares internacionales para la gestión de la información.
- c)** Garantizar la integridad y autenticidad de la información a preservar a largo plazo.
- d)** Garantizar la disponibilidad de la información durante el ciclo vital de los documentos.
- e)** Reducción progresiva del espacio físico que ocupa el almacenamiento de los archivos en papel y archivos en formato análogo.
- f)** Salvaguardar la información frente los cambios tecnológicos, evitando su obsolescencia durante el tiempo de retención estipulado.