ARCHIVE - INSIDE

Cálculo de la huella digital de documentos electrónicos

Departamento de Tecnologías Disruptivas
e Integridad de la Información

Equipo de Gestión Documental

Índice

1. Introducción 3
2. Obtención del contenido para calcular la huella 3
2.1 CASO A — XAdES y documentos de gran tamaño 4
2.2 CASO B — Formato no XML (PDF, PAdES, CSV…) 4
2.3 CASO C — Sin ValorBinario (CAdES detached/attached) 5
3. Traslado de la huella y el algoritmo al expediente 5

1. Introducción

Para garantizar la integridad de los documentos dentro de un expediente electrónico ENI, se calcula una huella digital (hash) de cada documento. Esta huella se añade al índice del expediente junto al identificador del documento y el algoritmo utilizado, permitiendo verificar que ningún documento ha sido manipulado.

> ⚠ **Nunca debe calcularse el hash del documento ENI completo. Solo debe calcularse sobre el fragmento específico indicado en cada caso.**

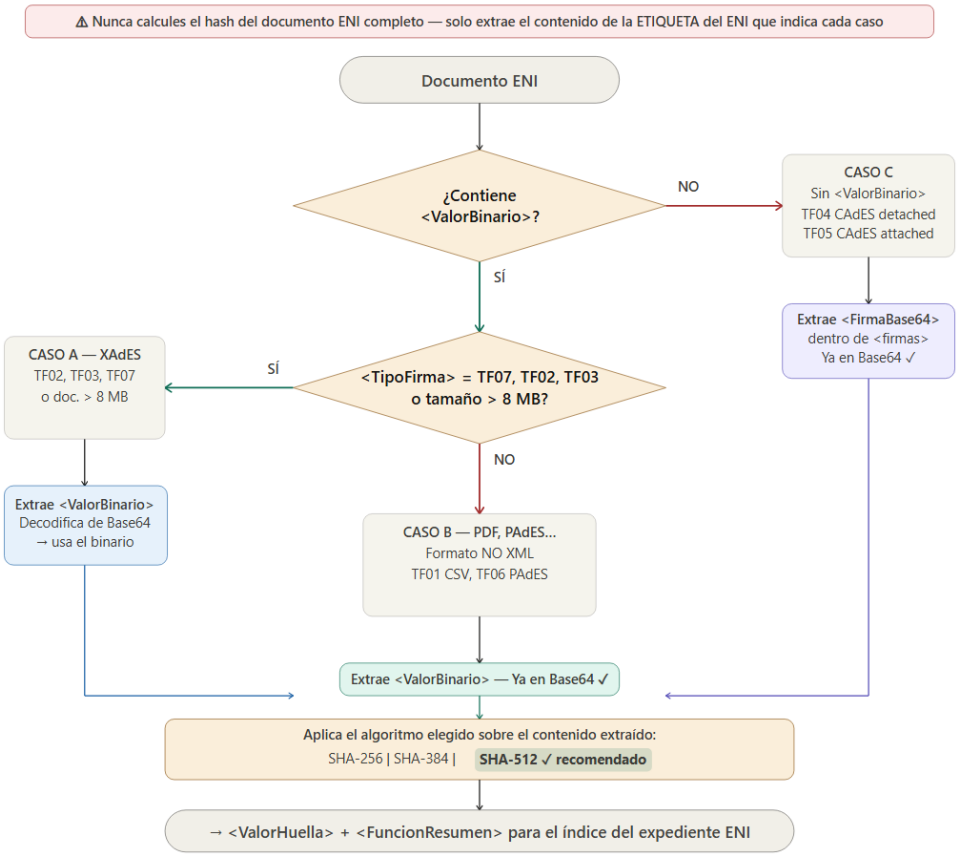
El contenido sobre el que calcular el hash varía según el tipo de firma del documento ENI. Los tres casos posibles se describen en la sección siguiente.

---

2. Obtención del contenido para calcular la huella

Dependiendo de la firma del documento ENI, el fragmento a utilizar para el cálculo varía, así como si debe estar codificado en Base64 o no.

El flujo de decisión es el siguiente:



CASO A	CASO B	CASO C
Tiene `<ValorBinario>` y TipoFirma = TF07, TF02, TF03 o documento > 8MB	Tiene `<ValorBinario>` y formato NO es XML (PDF, PAdES…)	NO tiene `<ValorBinario>` (CAdES TF04, TF05)
Extrae `<ValorBinario>` y **decodifica de Base64** → usa el binario resultante	Extrae `<ValorBinario>` y **úsalo tal cual**, ya viene en Base64	Busca `<FirmaBase64>` dentro de `<firmas>` y **úsalo tal cual**, ya en Base64

2.1 CASO A — XAdES y documentos de gran tamaño

Se aplica cuando el documento ENI contiene la etiqueta `<ValorBinario>` y además se cumple cualquiera de estas condiciones:

- `TipoFirma` = `TF07` (XAdES manifest)
- `TipoFirma` = `TF02` (XAdES internally detached signature)
- `TipoFirma` = `TF03` (XAdES enveloped signature)
- Tamaño del documento superior a 8 MB

CASO A — Contenido a utilizar para el hash	
Elemento	Etiqueta `<ValorBinario>` dentro de `<contenido>`
Tratamiento	Decodificar el valor de Base64 y **aplicar el hash al binario resultante**
Motivo	En estos formatos XAdES el contenido real del documento está encapsulado codificado en Base64 dentro del XML — hay que decodificarlo primero

> **i** La etiqueta `<ValorBinario>` se encuentra dentro de `<contenido Id="CONTENIDO_DOCUMENTO">` del documento ENI.

2.2 CASO B — Formato no XML (PDF, PAdES, CSV…)

Se aplica cuando el documento ENI contiene `<ValorBinario>`, no existe ningún `<TipoFirma>` con valor TF07/TF02/TF03, y el formato del documento **no es XML** (por ejemplo PDF, DOCX, CSV firmado…).

- `TF01` — CSV (firma tipo CSV)
- `TF06` — PAdES (PDF firmado)
- Cualquier otro formato binario no XML

CASO B — Contenido a utilizar para el hash	
Elemento	Etiqueta `<ValorBinario>` dentro de `<contenido>`
Tratamiento	Utilizar el valor directamente, **sin decodificar**, ya viene en Base64
Motivo	En estos formatos el `<ValorBinario>` ya contiene el fichero binario codificado en Base64 listo para calcular el hash

2.3 CASO C — Sin ValorBinario (CAdES detached/attached)

Se aplica cuando el documento ENI **no contiene** la etiqueta `<ValorBinario>`. En su lugar aparece una etiqueta `<referenciaFichero>` que apunta a la firma dentro de `<firmas>`.

- `TF04` — CAdES detached/explicit signature
- `TF05` — CAdES attached/implicit signature

CASO C — Contenido a utilizar para el hash	
Paso 1	Leer el valor de `<referenciaFichero>` (p. ej. `#FIRMA_0`)
Paso 2	Buscar en `<firmas>` la etiqueta `<firma Id="FIRMA_0">` (quitar el `#`)
Paso 3	Dentro de esa firma, localizar `<FirmaConCertificado>` → `<FirmaBase64>`
Tratamiento	Utilizar el valor de `<FirmaBase64>` directamente, **sin decodificar**, ya en Base64

> **i** El valor de `<referenciaFichero>` incluye el carácter `#` al inicio — hay que eliminarlo antes de buscar el Id de la firma.

---

3. Traslado de la huella y el algoritmo al expediente

Una vez obtenido el contenido correcto según el caso, se aplica el algoritmo de hash elegido. El resultado (huella digital) junto con el algoritmo utilizado se incluyen en el índice del expediente ENI.

Algoritmo	Valor de <FuncionResumen>	URI
SHA-256	`sha-256`	`http://www.w3.org/2001/04/xmlenc#sha256`

SHA-384	`sha-384`	`http://www.w3.org/2001/04/xmlenc#sha384`
SHA-512 ✓ recomendado	`sha-512`	`http://www.w3.org/2001/04/xmlenc#sha512`

En el índice del expediente ENI, junto al identificador ENI de cada documento, se debe informar:

- `<FuncionResumen>` — URI del algoritmo utilizado (ver tabla)
- `<ValorHuella>` — hash resultante de aplicar el algoritmo al contenido

> **i** Ambos valores deben incluirse por cada documento en el `<DocumentoIndizado>` del índice del expediente.