



UTE FORENSIA THOT

## F1.1.3. Modelo de Datos

**THOT**

Periodo de Informe 30/09/2025 a 27/02/2026

Fecha: 27/02/2026

Versión: 2.0

**Información de control del documento**

Descripción	Valor
Título del Documento:	Documento de Modelo de Datos
Nombre del Proyecto:	THOT
Autor del documento:	Sergio Zaera Mata, Sergio Queraltó Pereira, Jaime Castro Cernadas
Propietario del Proyecto:	UTE FORENSIA THOT
Director del Proyecto:	Roberto Gómez-Espinosa
Versión Doc.:	2.0
Confidencialidad:	Alta
Fecha:	27/02/2025

**Aprobación y revisión del documento:**

NOTA: Se requieren todas las aprobaciones. Se deben mantener registros de cada aprobación. Todos los revisores de la lista se consideran necesarios a menos que se indique explícitamente como Opcionales.

Nombre	Rol	Acción	Fecha
Sergio Zaera Mata	Jefe de Proyecto	Revisa	26/01/2026

**Historial de documentos:**

El Autor del Documento está autorizado a hacer los siguientes tipos de cambios al documento sin requerir que el documento sea aprobado nuevamente:

- Editorial, *formateo y ortografía*.
- Aclaración.

Para solicitar un cambio en este documento, póngase en contacto con el Autor o el Propietario del Documento.

Las modificaciones de este documento se resumen en la siguiente tabla en orden cronológico inverso (primero la última versión).

Revisión	Fecha	Creada por	Breve descripción de los cambios
0.0	07/10/25	Sergio Zaera Mata	Preparación ToC
0.1	03/11/25	Sergio Queraltó, Jaime Castro	Contribuciones técnicas iniciales
0.2	28/11/25	UTE ForensIA (Todos)	Revisión & Contribuciones adicionales
1.0	05/12/25	Sergio Zaera Mata	1º Borrador
1.1	16/01/26	UTE ForensIA (Todos)	Contribuciones técnicas
1.2	21/01/26	Sergio Queraltó, Jaime Castro	Revisión & Consolidación
2.0	26/01/26	Sergio Zaera Mata	2º Borrador

**ADVERTENCIA DE CONFIDENCIALIDAD Y RESPONSABILIDAD LEGAL**

Este documento contiene información confidencial y secretos empresariales propiedad de la UTE FORENSIA THOT, protegidos por la Ley 1/2019 de Secretos Empresariales, el artículo 13 de la Ley de Contratos del Sector Público (LCSP) y la Directiva (UE) 2016/943 sobre protección de know-how.

Se entrega exclusivamente para la finalidad prevista en el procedimiento administrativo o contractual.

Queda terminantemente prohibida su reproducción, divulgación, cesión o uso por terceros sin autorización expresa y por escrito.

El incumplimiento de estas obligaciones puede constituir:

- Infracción contractual, con las consecuencias previstas en la LCSP.
- Responsabilidad civil y penal, conforme a la Ley 1/2019 y al Código Penal (arts. 278 y ss.).
- Acciones judiciales inmediatas, incluyendo reclamación de daños y perjuicios y medidas cautelares.

Si usted no es el destinatario autorizado, debe comunicarlo de inmediato y proceder a la eliminación del documento. Cualquier uso indebido será perseguido con el máximo rigor legal”.

## Tabla de contenido

<b>1. RESUMEN EJECUTIVO.....</b>	<b>6</b>
<b>2. INTRODUCCIÓN.....</b>	<b>7</b>
2.1 Propósito del documento .....	7
2.2 Alcance del documento .....	7
<b>3 GRADO DE INNOVACIÓN .....</b>	<b>8</b>
<b>4 ESTRUCTURACIÓN Y FLUJOS DEL CONOCIMIENTO FORENSE .....</b>	<b>9</b>
4.1 Tipología y origen de los datos .....	9
4.2 Puntos de mejora en los flujos de información de PC.....	13
<b>5 DISEÑO ONTOLÓGICO .....</b>	<b>16</b>
5.1 Principios y beneficios .....	16
5.2 Estado del arte en el dominio forense .....	17
5.3 Dominios forenses considerados.....	18
5.4 Prototipo de Referencia para Cadena de Custodia y Análisis de Resultados.....	19
<b>6 MODELO DE DATOS PROPUESTO .....</b>	<b>26</b>
6.1 Alcance y convenciones .....	26
6.2 Entidades canónicas .....	26
6.3 Relaciones y cardinalidades .....	32
6.4 Diagramas entidad-relación (DER).....	34
6.5 Reglas de integridad .....	36
6.6 Principios seguidos .....	38
<b>7 DESTINO Y ALMACENAMIENTO DE LOS DATOS .....</b>	<b>40</b>
7.1 Objetivo y alcance de la persistencia.....	40
7.2 Principios de diseño para almacenamiento y custodia .....	40
7.3 Matriz de asignación de destino para los datos .....	41
7.4 Rutas y flujos de sincronización.....	42
<b>8 MANTENIMIENTO DE LA CADENA DE CUSTODIA .....</b>	<b>44</b>
8.1 Glosario de términos .....	44
8.2 Estrategia .....	44
<b>9 INTEROPERABILIDAD.....</b>	<b>46</b>
9.1 Interoperabilidad semántica y formatos de intercambio .....	46
9.2 Interoperabilidad entre repositorios de la plataforma THOT .....	47
9.3 Interoperabilidad entre bases de datos de PN .....	48

9.4	Interoperabilidad entre bases de datos externas .....	49
9.5	Interoperabilidad entre Lotes.....	50
<b>10</b>	<b>METODOLOGÍA, CALIDAD Y GOBIERNO DEL DATO .....</b>	<b>51</b>
10.1	Calidad del dato .....	51
10.2	Responsabilidades en el manejo de datos .....	52
10.3	Repositorios de confianza y principios FAIR .....	53
<b>11</b>	<b>MAPEO DE DATOS POR DISCIPLINA FORENSE.....</b>	<b>54</b>
<b>12</b>	<b>MAPEO DE DATOS POR SERVICIO .....</b>	<b>55</b>
<b>13</b>	<b>POLÍTICA DE DATOS Y SEGURIDAD DEL DATO .....</b>	<b>56</b>
13.1	Control de acceso y distribución.....	56
13.2	Anonimización de datos personales .....	57
13.3	Análisis de vulnerabilidades .....	57
13.4	Política de retención de los datos.....	58
<b>14</b>	<b>MIGRACIÓN Y EVOLUCIÓN DEL DATO .....</b>	<b>59</b>
14.1	Migración inicial.....	59
<b>15</b>	<b>ASPECTOS ÉTICOS .....</b>	<b>60</b>
15.1	Recogida y empleo de los datos .....	60
15.2	Sensibilidad, privacidad y protección de los derechos humanos .....	60
15.3	Responsabilidad, transparencia y supervisión .....	60
<b>16</b>	<b>CONCLUSIONES.....</b>	<b>62</b>
<b>ANEXOS.....</b>		<b>63</b>

## 1. RESUMEN EJECUTIVO

CONFIDENCIAL

## 2. INTRODUCCIÓN

### 2.1 Propósito del documento

Este documento establece el marco de referencia para el **Modelo de Datos** de la plataforma THOT a partir de un **prototipo ontológico básico** y deliberadamente acotado, con el fin de que se comprendan con precisión los **principios**, la **terminología operativa** y los **mecanismos de trazabilidad** que gobiernan la gestión de vestigios, eventos y análisis dentro del flujo de trabajo forense. La ontología se utiliza como capa semántica de partida para fijar significados, relaciones y responsabilidades, y para derivar un modelo canónico coherente con los requisitos de custodia, auditoría y explotación posterior.

En consecuencia, el objetivo del documento no consiste en agotar todas las entidades, atributos y reglas posibles del dominio, sino en **mostrar una base formal mínima y verificable** sobre la que se construye el modelo de datos y su evolución controlada. Esta aproximación permite discutir con Policía Científica, sobre artefactos concretos y trazables, qué conceptos son nucleares, qué restricciones se consideran críticas y en qué puntos del flujo de trabajo se aplican los controles de integridad y calidad.

Asimismo, se evita describir o prescribir **herramientas concretas** para cada repositorio de datos. La selección tecnológica y la topología de persistencia por componentes se documentan en el **entregable de arquitectura**, mientras que aquí se fija la lógica del dato: qué se registra, cómo se identifica, cómo se relaciona y cómo se verifica.

### 2.2 Alcance del documento

El alcance comprende la definición de un **modelo canónico** alineado con el diseño ontológico y orientado a soportar: (i) la gestión y reconstrucción de la **cadena de custodia** a partir de hechos auditables, (ii) la trazabilidad de **análisis** y resultados, (iii) la formalización de **linaje** entre vestigios originales y derivados, y (iv) la preparación del dato para servicios de consulta e interoperabilidad semántica. Este alcance se desarrolla sobre dominios funcionales del ámbito forense ya delimitados en el documento (asuntos, vestigios, personas, análisis y auditoría, entre otros), manteniendo consistencia terminológica y evolución controlada.

En términos de persistencia, el documento delimita qué se considera **fuentes de verdad** (registro canónico transaccional), qué repositorios actúan como **proyecciones** optimizadas para consulta y explotación, y qué registros constituyen **evidencia verificable** en el plano inmutable judicial, con separación explícita entre dato, metadato y prueba criptográfica.

Quedan fuera de alcance, por estar cubiertos en entregables específicos o por requerir validación operativa posterior, los detalles de implementación por producto (motores concretos, despliegue, parametrización y operación), así como el catálogo completo de integraciones y adaptadores con sistemas internos o externos. El documento se mantiene como referencia técnica para cimentar decisiones de diseño y para guiar la ampliación progresiva del modelo conforme se consolide la colaboración con Policía Científica.

### 3 GRADO DE INNOVACIÓN

Se añadirá en el próximo entregable

CONFIDENCIAL

## 4 ESTRUCTURACIÓN Y FLUJOS DEL CONOCIMIENTO FORENSE

### 4.1 Tipología y origen de los datos

El proyecto THOT integra múltiples casos de uso operativos y de apoyo a la actividad forense de la Policía Nacional, y abarca un ecosistema de datos extraordinariamente diverso, que contiene desde información administrativa, documental y procedimental hasta biometría (huellas, ADN, facial, dental), vestigios físicos y datos de escena, así como registros de calidad, inteligencia y gestión interna. Estos datos pueden proceder de un amplio abanico de fuentes y sistemas, entre los que se encuentran distintas bases de datos de gestión y biométricas, los registros de extranjería y asilo, los sistemas judiciales y penitenciarios, diversos repositorios europeos e internacionales (EURODAC, PRÜM, INTERPOL, SIS), así como las herramientas de inteligencia y los sistemas de recursos humanos y formación.

La identificación sistemática de todos estos tipos de datos y de sus orígenes es esencial para garantizar un tratamiento seguro, coherente y eficiente de la información en todo el ciclo de vida forense. Conocer qué datos existen, dónde se generan, en qué sistemas se almacenan y cómo se relacionan entre sí permite definir medidas de protección adecuadas (ENS, RGPD, AI Act), evitar duplicidades e inconsistencias, asegurar la trazabilidad y la cadena de custodia, diseñar integraciones fiables entre bases de datos heterogéneas y, en última instancia, disponer de una base sólida y auditada sobre la que apoyar tanto la generación de inteligencia como la toma de decisiones operativas y estratégicas.

#### 4.1.1 Tipologías principales de datos

A partir de un análisis identificativo, pueden distinguirse varias tipologías de datos claramente diferenciadas, que se describen a continuación:

- **Datos administrativos, de filiación y procedimentales.** En prácticamente todos los casos de uso se tratan datos personales y administrativos básicos: identidad de las personas (nombre, apellidos, filiación, fecha y lugar de nacimiento, nacionalidad), así como información de los procedimientos policiales y judiciales asociados (número de diligencias, tipo de hecho o evento migratorio, órgano judicial competente, fase procesal, estado de expedientes de extranjería, etc.).
- **Datos documentales y formularios.** El proyecto maneja un amplio conjunto de documentos estructurados y semiestructurados, tanto físicos como digitales, entre los que se incluyen:
  - Formularios y atestados policiales (Plan Nacional de Identificados, impresos V082/V094, actas de inspección técnico-policia, actas de levantamiento, denuncias de desaparición, formularios AM/PM INTERPOL, formularios de asilo y extranjería, oficios y solicitudes de análisis).
  - Informes periciales y técnicos emitidos por laboratorios y unidades de Policía Científica.
  - Documentación de calidad (Manual de Calidad, Procedimientos Generales y de Ensayo, registros y formatos) y documentación de formación (programas formativos, materiales docentes, resultados de pruebas).
- **Datos biométricos, biológicos y dentales.** Los datos biométricos se manejan como identificadores primarios o secundarios, y se relacionan con datos administrativos, de procedimiento y de caso en diversos casos de uso. Los datos biométricos abarcan:
  - Biometría lofoscópica: huellas dactilares (rodadas y planas), palmares y otras impresiones específicas. Se capturan en tinta o mediante dispositivos digitales (LiveScan, BlueCheck, CQG) y se almacenan y cotejan en los sistemas nacionales de biometría (ABIS/SAID) y, en su caso, en EURODAC o a través de PRÜM/INTERPOL.
  - Biometría facial: fotografías de reseña y de identificación (vistas de rostro y cuerpo), utilizadas tanto en contextos de reseña policial y extranjería como en identificación de víctimas.

- Datos biológicos/genéticos: perfiles de ADN obtenidos a partir de muestras biológicas recogidas en escenas, sobre personas detenidas o en contextos de identificación de víctimas. Se almacenan y gestionan a través de CODIS/SDIS y se usan tanto en ámbito nacional como en intercambios PRÜM/INTERPOL.
  - Datos dentales: fichas odontológicas, radiografías y documentación asociada, especialmente relevantes en el contexto de Identificación de Víctimas en Desastres.
- **Vestigios físicos, evidencias forenses y datos de escena.** Estos datos se recogen siguiendo procedimientos normalizados y se registran siguiendo una cadena de custodia estricta. Además, se relacionan con actividades de análisis en laboratorio y se integran en infografías forenses y reconstrucciones de hechos. La actividad de inspección técnico-policial y la gestión del ciclo de vida de vestigios implican la recogida y tratamiento de múltiples tipos de evidencias:
  - Muestras biológicas (ADN), fibras, pinturas, suelos, sustancias químicas, drogas, acelerantes y restos en general.
  - Armas de fuego, armas blancas, proyectiles y vainas, así como su información balística asociada.
  - Documentos físicos, dispositivos electrónicos, drones u otros equipos recuperados.
  - Información de fijación de escena (fotografía, vídeo, vídeo 360°, imágenes hiperespectrales, modelos 3D, planimetrías) así como la caracterización del entorno (zonificación, condiciones, hipótesis iniciales).
- **Metadatos de trazabilidad, cadena de custodia y aseguramiento de la calidad.** A lo largo de todos los casos de uso se generan metadatos orientados a garantizar la integridad, trazabilidad y auditabilidad de la información:
  - Identificadores únicos de reseñas, asuntos y vestigios.
  - Fechas, horas, localizaciones, operadores y unidades intervinientes en cada fase (captura de biometría, recogida de vestigios, traslado, análisis, emisión de informes, intercambios internacionales).
  - Registros de movimientos físicos (envíos entre plantillas y laboratorios, altas/bajas en almacenes) y reglas de negocio ligadas a la recepción y validación de vestigios.
  - Registros de calidad y auditoría (desviaciones, no conformidades, acciones correctivas, evidencias de auditorías internas y de ENAC).
  - Parámetros de laboratorio y datos técnicos de ensayo necesarios para demostrar conformidad con normas ISO 17020/17025/21043.
- **Datos de casos, inteligencia y explotación analítica.** En el ámbito de la inteligencia policial (Caso 7) y del seguimiento global del rendimiento forense (Casos 5, 6 y 8), se manejan:
  - Datos de casos e investigaciones (número de caso, tipo de delito, localización, estado, unidades implicadas).
  - Resultados de cotejo (matches positivos o técnicos entre vestigios, personas y armas, a nivel nacional e internacional).
  - Relaciones entre casos, vestigios, personas y armas, representadas en estructuras de grafos, cronologías o mapas.
  - Indicadores agregados y estadísticas (número de identificaciones, tiempos de respuesta, distribución geográfica, eficacia de técnicas, capacidad operativa).
- **Datos de gestión interna, formación, recursos humanos y medios técnicos.** Finalmente, el proyecto utiliza y genera datos de gestión interna vinculados a la actividad forense. Estos datos son esenciales para dimensionar, planificar y sostener la actividad pericial:
  - Datos de personal (estructura organizativa, roles, especialidades, rotación, jubilaciones).
  - Históricos de formación y cualificación (cursos, prácticas, pruebas de competencia, ejercicios ciegos e intercomparaciones).

- o Datos de medios técnicos e inventario (equipos, EPIs, fungibles, consumos, mantenimiento).
- o Registros administrativos de entrada y salida de documentos, materiales y vestigios en REGPOL y MINPOL.

#### 4.1.2 Origen y fuentes de los datos

Muchos de los datos descritos procederán de la propia **intervención operativa en la escena**. Sin embargo, THOT también debe poder integrar los datos almacenados en los diversos sistemas y repositorios ya existentes en el ecosistema de Policía Nacional. Teniendo en cuenta esta doble vertiente del proyecto, los principales orígenes pueden clasificarse en:

- Fuentes de datos vinculadas a la escena y otras actividades operativas.
  - o Inspección técnico-policial (ITP): punto de origen de vestigios físicos y biológicos, información de fijación (fotografía, vídeo, 360°, hiperespectral, modelos 3D), caracterización del entorno y planificación de la búsqueda.
  - o CIMACC / 091: centro de mando y control y sistema de gestión de llamadas donde se registran los avisos iniciales, la localización, el tipo de suceso, la hora y las unidades enviadas, constituyendo la primera fuente estructurada de información operativa sobre el incidente.
  - o Unidades de Seguridad Ciudadana y Policía Judicial: generan los partes e informes iniciales de actuación, con una descripción preliminar de los hechos, personas presentes, posibles testigos, incidencias y medidas adoptadas antes de la llegada de Policía Científica.
  - o Reseña en dependencias policiales: captura presencial de datos de filiación, biometría (huellas y fotografías de reseña), formularios oficiales (Plan Nacional de Identificados, impresos V082/V094, órdenes de reseña) y metadatos de control, directamente en comisaría en el momento de la detención o puesta a disposición.
  - o Operativos de interceptación y atención a migrantes y peticionarios de asilo: obtención in situ de datos personales, información del evento migratorio, documentación aportada, toma de huellas para identificación rápida y, en su caso, muestras de ADN para verificación de parentesco.
  - o Operativos de identificación de víctimas en desastres (IVD) y sucesos con víctimas múltiples: recogida en campo de datos Post Mortem (levantamiento, numeración de cadáveres y restos, pertenencias asociadas, necroimpresiones, muestras biológicas, datos dentales y rasgos identificativos) y de datos Ante Mortem preliminares, que luego se consolidan en PDyRH y sistemas asociados.
  - o Entrevistas y contacto con familiares y víctimas: generación de información cualitativa y cuantitativa (formularios AM, descripciones de cicatrices, tatuajes, prótesis, antecedentes médicos, objetos personales, circunstancias de la desaparición) que se registra inicialmente en papel o soportes digitales de campo antes de integrarse en los sistemas centrales.
  - o Dispositivos y herramientas de campo: captura primaria de datos mediante cámaras digitales, sistemas de vídeo 360°, equipos de imagen hiperespectral, dispositivos de captura dactilar portátiles (LiveScan, BlueCheck, CQG), tabletas y aplicaciones móviles, que actúan como origen tecnológico de muchos de los datos que posteriormente se estructuran en el resto del ecosistema.
- Sistemas policiales de identidad, denuncias y extranjería.
  - o PERSONAS: base nacional de reseñas y filiaciones, utilizada como referencia principal de identidad en reseñas de detenidos y en eventos migratorios.
  - o Sistemas de DNI/NIE/ADDEXTTRA y extranjería: verificación documental, asignación y gestión de NIE y expedientes asociados a asilo y cruce irregular.
  - o SIDENPOL y otros sistemas de denuncias: registro inicial de hechos delictivos y desapariciones.

- o MAPOL y sistemas geoespaciales: información cartográfica y de localización asociada a hechos y escenas.
- Bases de datos biométricas, genéticas y balísticas.
  - o ABIS/SAID: sistema nacional de biometría (huellas y, progresivamente, otras modalidades), origen y destino de datos lofoscópicos y faciales utilizados en reseñas, IVD, asilo y análisis de vestigios.
  - o CODIS/SDIS: repositorios y motores de búsqueda de perfiles genéticos, alimentados por laboratorios acreditados y usados tanto en comparaciones nacionales como en el marco PRÜM/INTERPOL.
  - o Sistemas balísticos (IBIS, EVOFINDER en otros países): soporte de la información balística de proyectiles y vainas, origen de los *matches* balísticos que se integran posteriormente en la inteligencia forense.
  - o EURODAC: base europea para huellas (y facial en evolución) de solicitantes de asilo y migrantes irregulares.
- Sistemas de gestión forense y de cadena de custodia.
  - o BINCIPOL: sistema central de gestión de asuntos, vestigios, entidades, actividades de análisis, envíos y adjuntos. Actualmente es la columna vertebral del ciclo de vida de datos forenses y el principal origen de información estructurada sobre vestigios y cadena de custodia.
  - o LIMS y bases especializadas de laboratorio (ADN, química, fibras, balística): origen de resultados técnicos y perfiles que luego se reflejan en CODIS, ABIS, IBIS y BINCIPOL.
- Sistemas judiciales, penitenciarios y de protección internacional.
  - o Sistemas judiciales y penitenciarios nacionales: receptores de documentación de reseña, ITP, informes y resultados, y origen de referencias procesales que se asocian a asuntos y vestigios.
  - o Sistemas de gestión de asilo y protección internacional, fiscalías de menores y servicios sociales: gestionan expedientes administrativos y judiciales de peticionarios de asilo, MENA y otros colectivos.
  - o Sistemas europeos e internacionales: SIS, VIS, ECRIS, así como los canales INTERPOL (I-24/7, SIRENE) y la red PRÜM, que actúan como fuente y destino de datos biométricos, nominales y de alerta.
- Sistemas de inteligencia y explotación analítica.
  - o GATI e Investiga: aplicaciones de inteligencia policial en las que se integran los datos forenses, vestigios anónimos y resultados de cotejo procedentes de BINCIPOL y de las BBDD biométricas/genéticas/balísticas.
  - o Herramientas analíticas como Analysis Notebook y PowerBI: actualmente se alimentan mediante exportaciones manuales para análisis de vínculos e informes estratégicos.
- Sistemas de gestión interna, personal y calidad.
  - o SIGESPOL y otros sistemas de RRHH: origen de datos de estructura organizativa, destinos y personal.
  - o Certool y herramientas de gestión de calidad: origen y repositorio de documentación de calidad, registros de auditoría, no conformidades y acciones correctivas.
  - o REGPOL y MINPOL: registros oficiales de entrada/salida de documentos, materiales y vestigios.
- Canales ofimáticos y soportes no estructurados.
  - o Además de las bases de datos y sistemas formales, una parte relevante de la información actualmente se genera y circula a través de correos electrónicos, ficheros Excel y otros documentos ofimáticos. Estos canales constituyen, de facto, fuentes de datos que el proyecto pretende estructurar e integrar, especialmente en lo relativo a solicitudes y resultados de intercambios internacionales (Caso 6), gestión de formación y medios (Caso 8) y explotación de BINCIPOL (Casos 5 y 7).

## 4.2 Puntos de mejora en los flujos de información de PC

Para ilustrar las oportunidades de mejora en los flujos de información de PC, se toma como referencia el caso de uso 8 “Gestión Interna, Formación y Medios Técnicos”. Se toma como ejemplo este caso de uso porque es uno de los más sencillos a nivel descriptivo, pero al mismo tiempo refleja muy bien la problemática general en muchos de los flujos de información de la Policía Científica. La Figura 1 muestra un diagrama que representa el flujo de información actual asociado a la gestión de recursos en Policía Científica, elaborado con datos y notas tomados de las Jornadas de Formación impartidas por Policía Científica. Este diagrama está organizado en cuatro niveles verticales. En la parte superior se sitúan los generadores de datos, que corresponden a distintas áreas funcionales (RRHH, Formación, Instructores de Formación, Calidad, Compras, Personal Operativo y Laboratorios de Análisis), junto con un actor potencial aún no consolidado para analítica (“Analistas de datos”). Desde estas áreas se originan distintos tipos de datos, agrupados por naturaleza: información de personal (efectivos, roles, especialidades, competencias, antigüedad), datos de formación (materiales, contenidos curriculares), datos de cualificación (resultados de pruebas, registros de asistencia y cómputos horarios), datos de calidad (certificaciones y normas), datos de inventario y stock, datos de consumo y necesidades (incluyendo alertas de fungibles y equipos), datos de registro y trazabilidad (documentos y/o metadatos vinculados a actuaciones), datos documentales (normativas, informes, facturas) y datos estadísticos.

En el siguiente nivel se muestran las **bases de datos y repositorios** donde esa información termina almacenándose o circulando: sistemas corporativos específicos (por ejemplo, para gestión de personal o certificación), junto con repositorios de uso transversal y no homogéneo como servidores de correo, archivos Excel y archivadores físicos en comisaría, además de plataformas operativas y documentales internas (por ejemplo, repositorios de registro/gestión documental). La presencia de múltiples flechas cruzadas indica que un mismo tipo de dato puede terminar en varios repositorios y que un mismo repositorio puede recibir información desde distintos generadores.

Finalmente, el nivel inferior identifica los **consultores de datos** (quienes consumen la información), entre los que aparecen RRHH, Formación, Instructores, Calidad, Compras, Analistas/Superiores y Personal Operativo, así como Auditores Externos. Las conexiones reflejan que la consulta no se realiza desde un único punto, sino a través de varios repositorios y con dependencias cruzadas entre áreas, lo que evidencia un ecosistema de intercambio heterogéneo en el que los flujos de información se ramifican, se recombinan y, en ciertos casos, vuelven a circular hacia actores distintos a los generadores originales.

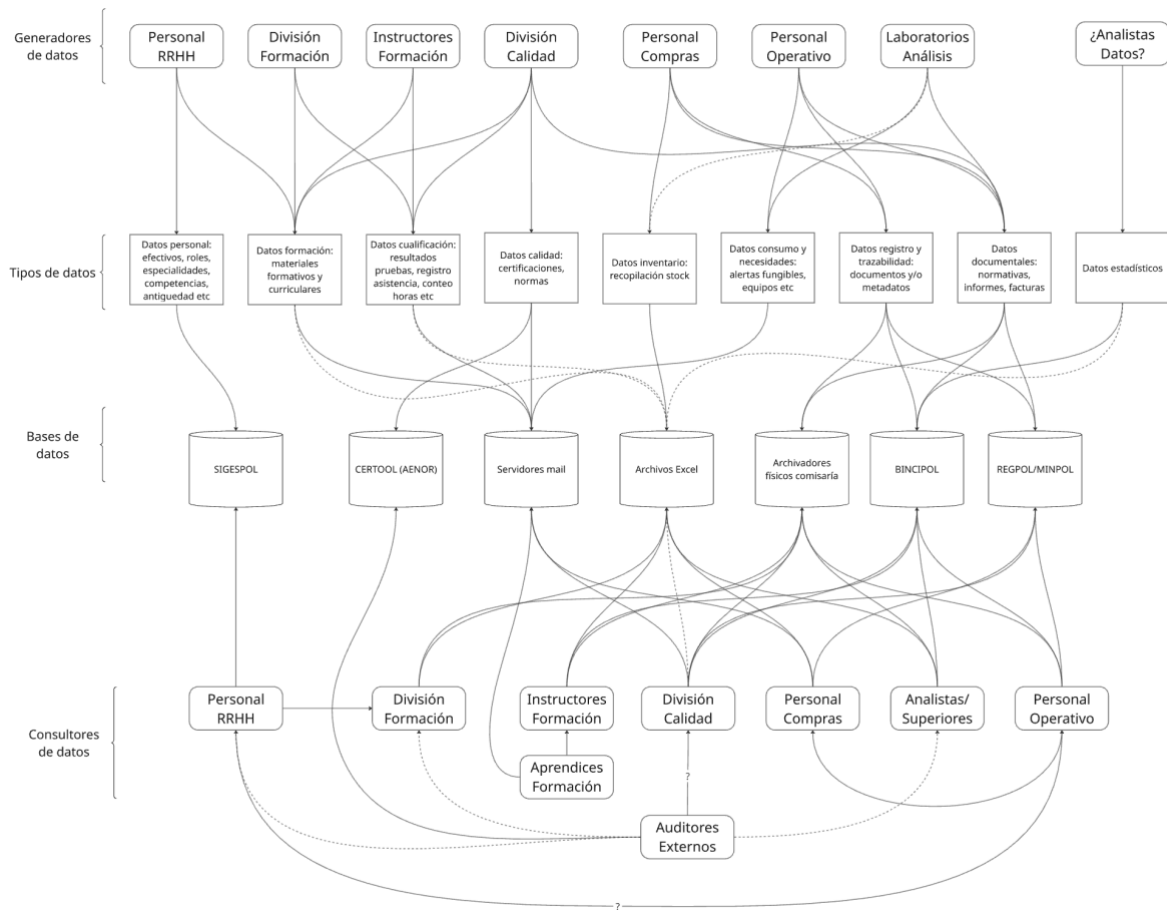


Figura 1 Flujo de información del caso de uso 8 “gestión interna, formación y medios técnicos” a modo de ejemplo

El flujo de información representado muestra una topología de tipo *many-to-many*: múltiples áreas generan información, dicha información se clasifica en varios tipos de datos y termina persistiendo en repositorios heterogéneos que, a su vez, son consumidos por múltiples perfiles. La presencia de conexiones cruzadas indica que el mismo dato circula por vías paralelas y, en ocasiones, regresa a unidades distintas a las generadoras. Este patrón incrementa la probabilidad de duplicidad de registros, latencia en la disponibilidad del dato, pérdida de trazabilidad y un coste elevado de coordinación entre áreas para reconciliar versiones y validar decisiones operativas.

Se identifican como puntos de fricción los repositorios que actúan como *hubs* informales, en particular los servidores de correo, los archivos Excel y los archivadores físicos. Al concentrar intercambios y consolidaciones, estos medios introducen dependencias críticas de personas concretas para producir o localizar la información vigente. Además, dificultan la determinación de la “versión válida” del dato cuando existen adjuntos replicados o hojas divergentes, y complican la auditoría y la reproducibilidad de los procesos al no disponer, por defecto, de control de versiones, trazabilidad de modificaciones ni vínculos semánticos explícitos con el resto del ecosistema.

En paralelo, se observa ausencia de una fuente de verdad definida por dominio funcional. Para ámbitos como personal/competencias, formación/cualificación, inventario/consumo y documentación, el mismo tipo de información se materializa en más de un repositorio y mediante canales de intercambio distintos. Esta fragmentación genera inconsistencias entre métricas según el origen consultado, obliga a

reconciliaciones manuales y desplaza decisiones operativas hacia el momento en que se “consolidan” datos, aumentando tiempos de respuesta y reduciendo la fiabilidad del seguimiento.

La trazabilidad documental aparece distribuida entre repositorios con naturalezas distintas, incluyendo almacenamiento físico. Cuando los datos de registro y trazabilidad (documentos y/o metadatos) y los datos documentales (normativas, informes, facturas) se reparten entre varias ubicaciones, se degrada la capacidad de demostrar, de forma sistemática, qué documento aplica a qué actividad, recurso o actuación. Además, la verificación de cumplimiento se vuelve dependiente de rutas informales (correo, carpetas personales, archivadores), lo que incrementa el esfuerzo de preparación ante auditorías y reduce la uniformidad en el acceso a la evidencia.

Otro punto de fricción se sitúa en la interfaz entre compras, operación y laboratorios, especialmente para los dominios de inventario/stock y de consumo/necesidades/alertas de fungibles y equipos. Cuando parte del circuito se soporta en Excel y correo, se introducen retardos en alertas, desalineación entre stock real y stock registrado y ausencia de un historial trazable de decisiones (por ejemplo, por qué una solicitud se ejecuta, se difiere o se descarta), dificultando la gestión preventiva y la planificación de recursos.

Finalmente, la función analítica se muestra como un componente no consolidado. La presencia de un nodo interrogativo para analítica y la separación del bloque de datos estadísticos sugieren que la explotación transversal se realiza de forma puntual y sin una función establecida. Este modo de operación reduce la continuidad en indicadores, impide la comparación sistemática en el tiempo y limita la capacidad de introducir mejoras sostenidas basadas en evidencias.

A partir de estas observaciones, se concluye que uno de los objetivos principales de THOT es clarificar fuentes de verdad por dominio y reducir los canales informales a un papel transitorio. Como criterio de diseño de THOT, se dispondrá de un repositorio canónico por dominio (personal/competencias, formación/cualificación, inventario/consumo y documentación), dejando Excel/correo/papel para transición o contingencia. Cuando el soporte físico se mantenga por obligación procedimental, su gestión deberá apoyarse en digitalización e indexación con metadatos mínimos.

Asimismo, otra de las prioridades de THOT será el enlazado y la normalización entre dominios mediante identificadores estables, adoptando IDs únicos para persona, recurso/equipo y documento, con versionado y estado. En los puntos críticos, la trazabilidad deberá consolidarse por eventos, registrando acciones relevantes como hechos auditables sin sobrescritura. La gestión documental deberá unificarse bajo control de metadatos, versionado y vigencia, y la explotación analítica deberá institucionalizarse para evitar silos, con responsabilidades claras e indicadores mínimos alimentados desde fuentes de verdad.

Estas conclusiones se incorporan como criterios de diseño de la ontología y del modelo de datos, dado que una ontología delimita significados, relaciones, identificadores y vocabularios controlados y permite estructurar el flujo de trabajo como hechos y vínculos explícitos, reforzando trazabilidad y reduciendo reconciliaciones manuales.

## 5 DISEÑO ONTOLÓGICO

### 5.1 Principios y beneficios

En el contexto de THOT, una ontología se entiende como una especificación explícita del significado de las entidades forenses relevantes y de sus relaciones, de forma que el sistema no solo almacena datos, sino que preserva su interpretación operativa. Esta capa semántica se sitúa por encima de la implementación física y se utiliza como referencia estable para derivar el modelo de datos canónico, definir contratos de integridad y garantizar coherencia entre servicios y repositorios. La ontología se construye sobre un conjunto de dominios funcionales previamente delimitados (asuntos, vestigios, personas, análisis, auditoría y trazabilidad, entre otros), de manera que el alcance queda acotado y evoluciona de forma controlada conforme se incorporan necesidades adicionales del flujo de trabajo.

El primer principio consiste en eliminar ambigüedad mediante tipado y relaciones explícitas. Cuando un vestigio, un evento o un análisis se representan como clases diferenciadas, con propiedades y vínculos definidos, se reduce la dependencia de convenciones locales o de texto libre para inferir significado. Por ejemplo, en lugar de registrar una "transferencia de vestigio" como campo de texto libre que podría interpretarse de múltiples formas, se modela como una relación estructurada Evento  $\leftarrow$  Envío  $\rightarrow$  Vestigio con atributos explícitos (origen, destino, *timestamp*, responsable), lo que elimina ambigüedad y facilita consultas automatizadas. Este enfoque mejora la consistencia del registro entre unidades y laboratorios y permite que la misma entidad (por ejemplo, un vestigio) se interprete de forma uniforme a lo largo de todo su ciclo de vida, incluyendo su asociación con uno o varios asuntos cuando exista base procedimental.

Un segundo principio es la trazabilidad reconstruible como propiedad del modelo, no como efecto colateral. La ontología prioriza representar "hechos" y "vínculos" (qué acción ocurre, sobre qué entidad, por quién y cuándo) de manera que la cadena de custodia y el linaje de resultados se obtienen como reconstrucción verificable a partir de registros auditables. Esto facilita auditorías internas, revisiones de calidad y análisis retrospectivo de actuaciones, y prepara el terreno para separar, en secciones posteriores, la fuente de verdad operativa de las proyecciones orientadas a consulta y de la evidencia verificable en el plano inmutable.

Un tercer principio es la gobernanza mediante vocabularios controlados y reglas semánticas. La ontología aporta un lenguaje común para clasificar tipos de evento, tipos de vestigio, estados de análisis o escalas de calidad/confianza, lo que permite comparar y explotar información de forma agregada sin perder interpretabilidad. Cuando estos elementos se formalizan como dominios controlados y se acompañan de restricciones (por ejemplo, cardinalidades o coherencia entre relaciones), se reduce el riesgo de datos parcialmente válidos que dificulten la auditoría o la explotación analítica.

Desde el punto de vista de beneficios, el principal impacto recae en la optimización de la gestión forense en tres planos complementarios. En el plano operativo, se mejora la calidad del dato y la reproducibilidad del flujo de trabajo: las altas, transferencias, custodias, análisis y derivaciones quedan descritos con estructura homogénea, lo que reduce inconsistencias entre registros y simplifica controles de integridad en los puntos críticos de la operación. En el plano analítico, se habilitan consultas integrales que cruzan asuntos, vestigios, eventos, análisis, resultados y afirmaciones de vinculación sin necesidad de integraciones ad hoc por cada servicio, al estar las relaciones modeladas como parte del núcleo semántico. En el plano probatorio, la explicitación de procedencia y auditoría facilita establecer una separación clara entre operación y verificación: los eventos y sus huellas criptográficas se tratan como evidencia verificable cuando proceda, mientras que el resto del ecosistema se mantiene orientado a explotación y eficiencia.

Adicionalmente, una ontología bien definida mejora la interoperabilidad interna y externa sin exigir homogeneidad tecnológica. Al mantener el significado del dominio independiente de cómo se materializa (tablas relacionales, proyecciones a grafo, índices vectoriales o flujos de eventos), se preservan invariantes semánticos —identificadores, tipos y vínculos— que pueden serializarse en formatos de intercambio y alinearse con modelos estándar del sector. Esta base semántica, por tanto, prepara el documento para las secciones posteriores dedicadas al modelo de datos, al destino y almacenamiento de la información, y a la interoperabilidad semántica, evitando que la coherencia dependa de una única base de datos o de un único servicio.

## 5.2 Estado del arte en el dominio forense

En el dominio forense, la estandarización semántica se apoya en dos corrientes complementarias: (i) modelos ontológicos para describir evidencias, acciones y actores con significado formal y (ii) normas y especificaciones de procedimiento que fijan definiciones y requisitos para garantizar integridad, trazabilidad y aceptación probatoria. En conjunto, estos marcos permiten representar “qué ocurrió” (hechos y transferencias), “con qué” (ítems y recursos), “quién” (roles y responsabilidades) y “en qué condiciones” (protocolos, acreditación y calidad), preservando el linaje de la información a lo largo del flujo de trabajo.

En el ámbito de ciber-investigación y forense digital, UCO (*Unified Cyber Ontology*) proporciona una base conceptual y reutilizable para representar entidades del ecosistema de seguridad (objetos, identidades, acciones y contexto) de forma consistente entre herramientas y organizaciones. Sobre esa base, CASE (*Cyber-investigation Analysis Standard Expression*) formaliza el intercambio de información de investigación y facilita la representación del ciclo de vida de la investigación, incluyendo aspectos de procedencia y cadena de custodia en términos de “quién hizo qué, cuándo y dónde”. Estas ontologías priorizan la interoperabilidad y la combinación automatizada de información procedente de múltiples fuentes, resultando especialmente útiles cuando se necesita correlación entre evidencias, acciones y resultados a nivel de investigación técnica.

Para describir linaje de forma transversal al dominio (tanto humano como algorítmico), PROV-O constituye un estándar ampliamente adoptado para modelar procedencia mediante las nociones de entidad, actividad y agente, así como sus relaciones de generación, uso y atribución. PROV-O resulta adecuado para representar dependencias entre artefactos (por ejemplo, resultados derivados de un análisis) y para documentar la trazabilidad de transformaciones sin imponer un modelo forense específico, dado que está diseñado para especializarse por dominio cuando se requiere mayor precisión semántica.

Junto a los marcos ontológicos, el estado del arte incorpora estándares operativos que fijan definiciones y guías de actuación relevantes para la custodia e integridad. En evidencia digital, ISO/IEC 27037 establece directrices para identificación, recogida, adquisición y preservación de evidencias digitales; además, ISO/IEC 27041 e ISO/IEC 27043 abordan, respectivamente, el diseño/adequación de métodos de investigación y guías de proceso de investigación en contextos con evidencia digital. En el plano conceptual forense general, ISO 21043-1 normaliza terminología del proceso forense “de la escena al tribunal” y sitúa explícitamente la cadena de custodia como elemento asociado a la integridad del ítem, contribuyendo a la consistencia terminológica entre unidades operativas y laboratorios. Como referencia definicional adicional, NIST recoge la cadena de custodia como registro cronológico de transferencia, manejo y almacenamiento de un ítem desde su recogida hasta su devolución o disposición final, reforzando el enfoque de secuencia verificable.

Finalmente, para interoperabilidad en intercambio de información entre organizaciones del ámbito de justicia y seguridad, se dispone de marcos como NIEM, orientado a un entendimiento semántico común de los datos intercambiados y a la estandarización de contenidos y procesos de intercambio. En ciberinteligencia, especificaciones como STIX 2.1 normalizan la representación y el intercambio de información de amenazas y

observables, resultando relevantes cuando los resultados analíticos o vínculos de inteligencia requieran alineación con formatos usados en intercambio interinstitucional.

Este conjunto de referencias consolida tres conclusiones operativas para el diseño que se desarrolla en las secciones siguientes. Primero, la trazabilidad robusta se expresa con mayor fiabilidad como hechos (eventos) y relaciones explícitas, complementadas por procedencia formalizable y definiciones operativas. Segundo, los modelos ontológicos de ciberinvestigación aportan patrones reutilizables de representación de acciones, actores y custodia que se integran con un modelo de vestigio y análisis cuando se requiere trazabilidad extremo a extremo. Tercero, se mantiene la necesidad de un prototipo de referencia acotado, alineado con la terminología operativa de Policía Científica y preparado para mapearse a estándares semánticos y a validación por restricciones, garantizando evolución controlada del modelo conforme avance la colaboración técnico-operativa.

### 5.3 Dominios forenses considerados

Para el diseño ontológico del modelo de datos de la plataforma THOT se tienen en cuenta los dominios funcionales recogidos en la Tabla 1. Cabe destacar que estos dominios se listan como parte del modelo completo previsto, pero en el prototipo de referencia presentado en este documento, sección 5.4, se hace un modelado más limitado que, por ejemplo, no explicita los dominios 'Lugares y Escenas' o 'Configuración y Maestros' a nivel de entidades. En el caso del primero de estos dominios se mantiene 'Ubicación' como atributo por motivos de simplicidad. La incorporación de entidades completas dentro de estos dominios, con sus relaciones y especializaciones, se realizará en fases posteriores en cooperación con Policía Científica.

Dominio	Descripción	Entidades principales
Asuntos	Expedientes forenses, investigaciones, relaciones	Asunto, Expediente, Incidente
Evidencias y Vestigios	Objetos físicos y digitales recogidos en escena	Vestigio, Evidencia, Muestra, Archivo Digital
Personas	Sujetos implicados en investigaciones	Agente, Civil, Detenido, Víctima, Testigo, Analista
Lugares y Escenas	Ubicaciones geográficas y escenas del delito	Escena, Lugar, Vehículo, Ubicación
Análisis y Resultados	Outputs de laboratorio y modelos IA	Resultado, Informe Pericial, Hipótesis, Score IA, Vinculación
Operaciones y Flujos	Ejecución de procesos y tareas	Tarea, Proceso, Asignación, Solicitud, Diligencia, Alerta
Configuración y Maestros	Catálogos y configuración del sistema	Usuario, Rol, Permiso, Catálogo, Plantilla
Auditoría y Trazabilidad	Registro de todas las operaciones	Evento, Log, Firma, Sello de Tiempo, Hash

Tabla 1 Dominios funcionales del diseño ontológico

## 5.4 Prototipo de Referencia para Cadena de Custodia y Análisis de Resultados

### 5.4.1 Descripción

El Prototipo de Referencia para Cadena de Custodia y Análisis de Vestigios Figura 1) se define como una ontología inicial orientada a estructurar, con un primer nivel de formalización, los elementos nucleares que intervienen en la trazabilidad forense de vestigios. Se modelan las entidades principales del dominio, sus especializaciones y las relaciones que permiten describir el ciclo de vida operativo del vestigio desde su incorporación a custodia policial hasta la generación de resultados analíticos y la formulación de afirmaciones de vinculación. La ontología se presenta como un artefacto de referencia para el diseño del modelo de datos, de modo que las estructuras lógicas del dominio queden explícitas y reutilizables en el resto del sistema.

En el núcleo del prototipo se sitúa la clase abstracta 'Vestigio', definida mediante algunos atributos descriptivos ('Tipo', 'Naturaleza', 'Carácter', 'Descripción') y especializada en 'VestigioOriginal' y 'SubVestigio'. Esta distinción separa, de forma explícita, los vestigios incorporados al sistema por eventos de entrada (ITP, autopsia, reseña etc.) de aquellos que se derivan como producto de procesos analíticos, lo que facilita la trazabilidad de transformaciones y la gestión del linaje de evidencias. Los vestigios se vinculan con 'Asunto' mediante una relación de pertenencia de cardinalidad múltiple, lo que permite representar escenarios en los que un vestigio se asocia a más de un contexto de investigación cuando existe una base procedimental para ello. La dinámica operativa se articula mediante la clase abstracta 'Evento', que agrupa acciones temporales ('FechaInicio', 'FechaFin') y se especializa en 'Entrada', 'Envío', 'Almacenamiento' y 'Análisis'. Sobre esta base se representa que un vestigio sufre eventos a lo largo de su vida, de manera que se mantiene un historial de acciones aplicadas a un vestigio. En particular, los eventos 'Entrada' (con subtipos como 'ITP' y 'Autopsia') se conectan con la creación de 'VestigioOriginal', mientras que los eventos 'Análisis' se conectan con la generación de 'SubVestigio' y con la producción de 'Resultado'.

El prototipo incorpora además entidades y relaciones necesarias para describir el análisis forense con criterios de validez operativa. Un 'Análisis' se ejecuta por uno o varios 'Analista' y sigue uno o varios 'Protocolo', representando explícitamente que una ejecución analítica puede apoyarse en procedimientos compuestos. Para caracterizar la ejecución se contemplan recursos consumidos o utilizados, mediante relaciones con 'Reactivo' y 'Equipamiento'. El compromiso de calidad se incorpora mediante 'Acreditación', de forma que un análisis puede obtener diversas acreditaciones y los protocolos pueden habilitar distintas acreditaciones. La relación entre protocolos y competencias se formaliza con 'Especialidad' y sus subtipos (por ejemplo, 'AnálisisADN' y 'Balística'), de forma que se expresa que el protocolo requiere especialidades y que el analista posee especialidades. Finalmente, la transición desde resultados técnicos a conclusiones operativas o judiciales se modela mediante 'AfirmaciónVinculación', sustentada por uno o varios resultados y asociada a uno o varios vestigios y a un único 'Civil' (clase diseñada para acoger distintas filiaciones, tales como Investigado, Detenido, Testigo etc.), incorporando además atributos de calidad y confianza en 'Resultado' y 'AfirmaciónVinculación' respectivamente. Esta estructura permite separar el dato observado (resultado), la trazabilidad (vestigios y eventos) y la interpretación (afirmación) sin colapsar conceptos heterogéneos en una misma entidad.

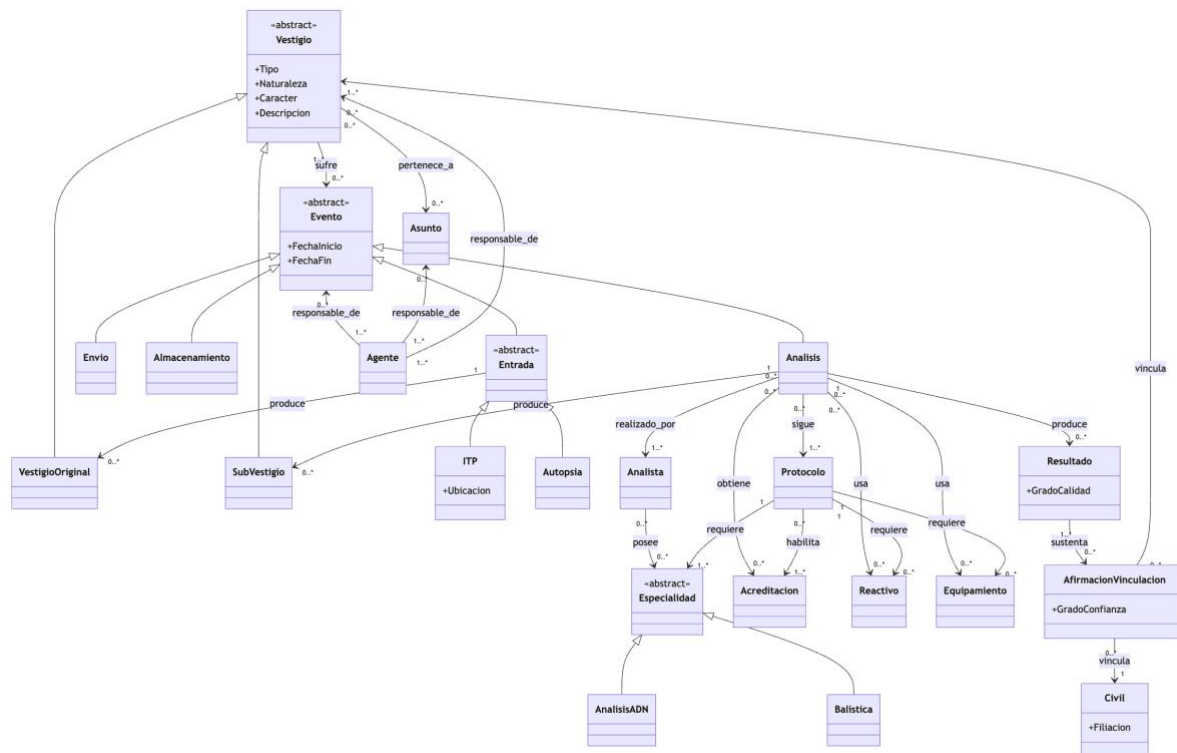


Figura 2 Prototipo de Referencia para Cadena de Custodia y Análisis de Vestigios

#### 5.4.2 Fortalezas

Las fortalezas del prototipo se derivan de su orientación explícita a la trazabilidad y a la separación de responsabilidades semánticas. La división entre vestigio original y subvestigio aporta una base clara para representar derivaciones, submuestras o productos intermedios generados en laboratorio, evitando que la creación de nuevas evidencias se confunda con la incorporación inicial al caso. La jerarquía de eventos permite describir de manera uniforme acciones diversas sobre el vestigio y, al mismo tiempo, mantener especializaciones para capturar particularidades operativas. La inclusión de protocolos, recursos, especialidades y acreditaciones proporciona un punto de partida para conectar el “qué se hace” con el “cómo se hace” y el “en qué condiciones se considera válido”, sin necesidad de introducir desde el primer momento detalles instrumentales excesivos. La separación entre resultado y afirmación de vinculación mantiene una frontera lógica entre medición y conclusión, lo que reduce ambigüedades y facilita el control posterior de calidad, revisión y auditoría.

#### 5.4.3 Limitaciones

Las limitaciones del prototipo se asumen como propias de una ontología en fase preliminar. La ontología no codifica todavía, de forma completa, todos los conceptos operativos que intervienen en el procedimiento forense ni las variaciones organizativas y normativas que pueden darse entre unidades o laboratorios. Por citar un ejemplo, se ha evitado conscientemente crear la entidad 'Lugar' para vincularla a la entidad 'ITP' y se ha establecido 'Ubicación' como un simple atributo de esta entidad por motivos de simplicidad y claridad del prototipo. Por otro lado, se ha omitido la declaración del término forense 'TipoEntidad' (Persona, Vehículo o Lugar) como atributo de la entidad 'Entrada' a pesar de que 'Entidad' es un atributo con un impacto importante en la labor actual de trazabilidad forense. Esta decisión se ha tomado para evitar confusiones con el término "entidad", que es mucho más amplio y abstracto en el dominio ontológico. Sin embargo, estas decisiones solo atañen al presente documento y no tendrán mayor repercusión en el producto final para Policía Científica. De hecho, la evolución de este modelo se realizará mediante refinamiento iterativo y validación continua con Policía, integrando criterios operativos, terminología institucional, excepciones procedimentales y necesidades de explotación analítica. En consecuencia, el prototipo se tratará como un componente vivo del “Modelo de

datos” y su estabilidad semántica se consolidará conforme se disponga de casos de uso, escenarios de prueba y retroalimentación operacional verificable.

#### 5.4.4 Reglas y restricciones semánticas

Para que el prototipo mantenga coherencia interna y evite la aparición de estados inconsistentes durante su implementación, tanto este como cualquier diagrama ontológico debe acompañarse de reglas y restricciones semánticas. Estas restricciones no se infieren de manera automática a partir de las relaciones y cardinalidades del diagrama, por lo que se expresarán de forma explícita como invariantes del modelo. A modo de base, se consideran relevantes, entre otras, las siguientes:

- La consistencia temporal de todo evento, imponiendo que 'FechaFin' no anteceda a 'FechaInicio'
- La coherencia de acreditación, exigiendo que toda acreditación obtenida por un análisis se encuentre habilitada por al menos uno de los protocolos seguidos por ese análisis
- El cumplimiento de requisitos de protocolo, imponiendo que las especialidades requeridas por los protocolos estén cubiertas por el conjunto de analistas que realizan el análisis, y que los reactivos y equipamientos requeridos se encuentren presentes entre los recursos usados
- La coherencia de vinculación, imponiendo que toda afirmación de vinculación se sustente en resultados existentes y que el grado de confianza se ajuste a un dominio controlado
- La consistencia de pertenencia a asunto en derivaciones, imponiendo que un subvestigio herede el contexto de asunto del vestigio del que deriva cuando se establezca trazabilidad de procedencia.

Estas reglas se implementarán como validaciones en capa de dominio, restricciones en el repositorio semántico cuando aplique, y controles de integridad en persistencia para evitar que el sistema acepte datos que contradigan la lógica operativa.

#### 5.4.5 Desarrollo bajo el estándar OWL2

El paso del prototipo preliminar al estándar OWL 2 se realizará mediante una traducción sistemática de clases, jerarquías y relaciones a constructos ontológicos normalizados. Las clases del prototipo se mapearán a owl:Class, mientras que las relaciones se materializarán como owl:ObjectProperty (y, cuando corresponda, owl:DatatypeProperty para atributos literales) con dominios y rangos explícitos. De este modo se fija, de forma verificable, qué tipos de entidades conecta cada propiedad. Por ejemplo, la propiedad ‘:sufre’ se define con dominio ‘:Vestigio’ y rango ‘:Evento’, mientras que ‘:sigue’ se define con dominio ‘:Análisis’ y rango ‘:Protocolo’. Las jerarquías de especialización se expresarán mediante rdfs:subClassOf, de forma que, por ejemplo, ‘:VestigioOriginal’ y ‘:SubVestigio’ queden como subclases de ‘:Vestigio’, y ‘:Entrada’ y ‘:Análisis’ como subclases de ‘:Evento’.

Además de la estructura taxonómica, el estándar permite introducir restricciones formales sobre relaciones y clases. Cuando el prototipo requiera límites cuantitativos, se utilizarán restricciones de cardinalidad (por ejemplo, owl:minCardinality, owl:maxCardinality y owl:cardinality) para reflejar condiciones del tipo “al menos uno”, “como máximo uno” o “exactamente uno” en determinadas propiedades. Estas restricciones se declaran axiomas de clase y pasan a formar parte de la semántica de la ontología, lo que posibilita detectar inconsistencias cuando los datos no cumplen dichas condiciones.

Las restricciones semánticas se formalizarán, en la medida de lo posible, mediante axiomas OWL 2. Un axioma OWL 2 es una declaración lógica que establece hechos o reglas declarativas sobre clases y propiedades (por ejemplo, subclases, dominios y rangos, disyunciones, equivalencias, restricciones de valores y cardinalidades). Su finalidad no es ejecutar validaciones procedimentales, sino habilitar razonamiento automático: un razonador

puede clasificar instancias, inferir pertenencias a clases o detectar que una base de conocimiento resulta inconsistente bajo los axiomas declarados. No obstante, se mantendrá la ontología dentro del perfil OWL 2 DL siempre que sea posible. OWL 2 DL es el subconjunto de OWL diseñado para equilibrar expresividad y garantías de razonamiento: permite restricciones ricas (disyunciones, cardinalidades, cuantificadores, etc.) manteniendo la decidibilidad, lo que implica que los procesos de inferencia y verificación terminan y son tratables con herramientas estándar. Cuando se requieran restricciones operativas que excedan lo expresable de forma natural en OWL 2 DL —por ejemplo, dependencias entre conjuntos de propiedades, validaciones cruzadas con condiciones, o comprobaciones sobre datos literales— se recurrirá a mecanismos complementarios.

En cuanto a la gestión de integridad de datos, el *payload* se canonicaliza (es decir, se normaliza a una representación estándar independiente del orden de campos, formatos de fecha o variaciones de serialización) antes de calcular el hash criptográfico, lo que garantiza que huellas idénticas correspondan a datos semánticamente equivalentes y permite verificación repetible.

En particular, se utilizarán dos familias de restricciones OWL frecuentes en modelado: las restricciones de tipo “al menos” (existenciales, owl:someValuesFrom) y de tipo “solo” (universales, owl:allValuesFrom). Las primeras expresan que existe al menos un valor de una propiedad que pertenece a una clase dada (por ejemplo, “todo Analisis sigue al menos un Protocolo”). Las segundas restringen los posibles valores de una propiedad a un conjunto de clases (por ejemplo, “todo lo que un Analisis sigue debe ser un Protocolo”). Estas restricciones son declarativas y permiten inferencia, pero no cubren con facilidad ciertas comprobaciones de integridad orientadas a datos, especialmente cuando se requiere contrastar valores entre propiedades o validar condiciones de negocio.

Para cubrir dichas necesidades se empleará SHACL (*Shapes Constraint Language*) como mecanismo de validación de datos RDF. SHACL define *shapes* o “perfiles de forma” que especifican restricciones verificables sobre los datos (cardinalidades, tipos, patrones, comparaciones, dependencias, valores permitidos), y se utiliza para comprobar de manera determinista si una instancia cumple o incumple las reglas definidas. Mientras OWL se centra en semántica e inferencia, SHACL se utiliza como validación y control de calidad de datos, de forma especialmente adecuada para invariantes operativas como consistencia temporal, obligatoriedad de campos o coherencia entre acreditaciones y protocolos seguidos.

A continuación, se incluyen ejemplos ilustrativos, en sintaxis Turtle para OWL y SHACL, que muestran cómo podrían declararse algunos elementos del prototipo y algunas restricciones representativas:

```
@prefix : <http://example.org/forense#> .
```

```
@prefix owl: <http://www.w3.org/2002/07/owl#> .
```

```
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
```

```
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
```

```
### Clases principales
```

```
:Vestigio a owl:Class .
```

```
:Evento a owl:Class .
```

```
:Analisis a owl:Class ; rdfs:subClassOf :Evento .
```

```
:Protocolo a owl:Class .
```

```
:Acreditacion a owl:Class .
```

### ### Propiedades (dominio y rango)

:sufre a owl:ObjectProperty ;

rdfs:domain :Vestigio ;

rdfs:range :Evento .

:sigue a owl:ObjectProperty ;

rdfs:domain :Análisis ;

rdfs:range :Protocolo .

:habilita a owl:ObjectProperty ;

rdfs:domain :Protocolo ;

rdfs:range :Acreditacion .

:obtiene a owl:ObjectProperty ;

rdfs:domain :Análisis ;

rdfs:range :Acreditacion .

### ### Restricciones OWL: cardinalidad y cuantificadores

:Análisis rdfs:subClassOf

[ a owl:Restriction ; owl:onProperty :sigue ; owl:minCardinality "1"^^xsd:nonNegativeInteger ],

[ a owl:Restriction ; owl:onProperty :sigue ; owl:allValuesFrom :Protocolo ] .

En el ejemplo anterior se declaran axiomas OWL que garantizan, por un lado, que todo análisis sigue al menos un protocolo (owl:minCardinality 1) y, por otro, que los valores de ‘:sigue’ sean siempre protocolos (owl:allValuesFrom). Estos axiomas habilitan razonamiento y detección de inconsistencias, pero no capturan por sí mismos dependencias del tipo “si se obtiene una acreditación, esta debe estar habilitada por algún protocolo seguido”, ya que esta condición es una validación cruzada que se expresa más directamente como restricción de datos.

El siguiente ejemplo muestra cómo se validan restricciones operativas mediante SHACL. Se definen *shapes* aplicables a nodos de tipo ‘:Evento’ y ‘:Análisis’. En el caso del evento, se valida que la fecha de fin no preceda a la de inicio. En el caso del análisis, se valida que exista al menos un protocolo seguido y que, si existe acreditación, esta esté habilitada por algún protocolo seguido.

@prefix : <http://example.org/forense#> .

@prefix sh: <http://www.w3.org/ns/shacl#> .

@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

:EventoShape a sh:NodeShape ;

```
sh:targetClass :Evento ;
```

```
sh:property [
  sh:path :FechaInicio ;
  sh:datatype xsd:dateTime ;
  sh:minCount 1 ;
];
```

```
sh:property [
  sh:path :FechaFin ;
  sh:datatype xsd:dateTime ;
  sh:maxCount 1 ;
];
```

```
# FechaFin >= FechaInicio (comparación entre propiedades)
```

```
sh:sparql [
  a sh:SPARQLConstraint ;
  sh:message "FechaFin no puede ser anterior a FechaInicio." ;
  sh:select """
    SELECT $this
    WHERE {
      $this :FechaInicio ?ini .
      $this :FechaFin ?fin .
      FILTER (?fin < ?ini)
    }
    """ ;
].
```

```
:AnalisisShape a sh:NodeShape ;
```

```
sh:targetClass :Analisis ;
```

```
sh:property [
  sh:path :sigue ;
  sh:minCount 1 ;
  sh:class :Protocolo ;
];
```

```
sh:property [
  sh:path :obtiene ;
```

```

sh:maxCount 1 ;
sh:class :Acreditacion ;
] ;

```

# Si existe acreditación, debe estar habilitada por algún protocolo seguido

```

sh:sparql [
  a sh:SPARQLConstraint ;
  sh:message "La acreditación obtenida debe estar habilitada por algún protocolo seguido." ;
  sh:select """
    SELECT $this
    WHERE {
      $this :obtiene ?acc .
      FILTER NOT EXISTS {
        $this :sigue ?p .
        ?p :habilita ?acc .
      }
    }
  """ ;
] .

```

Estos ejemplos muestran el reparto de responsabilidades entre OWL 2 y SHACL mientras que OWL 2 formaliza la semántica estructural (clases, propiedades, jerarquías, dominios/rangos y restricciones declarativas que facilitan inferencia), SHACL valida invariantes operativas y restricciones cruzadas que deben cumplirse en los datos para considerarlos consistentes con la lógica del dominio. En conjunto, la conversión a OWL 2 y la validación con SHACL permitirán disponer de un modelo interoperable, verificable y alineado con prácticas estándar en integración semántica y gobierno del conocimiento, manteniendo un control explícito sobre la consistencia y la calidad de los datos instanciados en el sistema.

## 6 MODELO DE DATOS PROPUESTO

La presente sección deriva directamente del Diseño ontológico descrito previamente y lo traduce a un modelo de datos coherente con los dominios funcionales y el Prototipo de Referencia para Cadena de Custodia y Análisis de Vestigios. Se define un núcleo canónico de entidades y relaciones que soporta trazabilidad, auditoría y explotación operativa en flujos de trabajo forenses, manteniendo independencia respecto de tecnologías concretas de persistencia.

### 6.1 Alcance y convenciones

El modelo de datos propuesto cubre el núcleo operativo del flujo de trabajo: gestión de Asunto, gestión y trazabilidad de Vestigio, registro de Evento como unidad de auditoría y cadena de custodia, y ejecución de Análisis como evento especializado que produce resultados y, en su caso, derivaciones. Esta base se completa con una entidad abstracta Persona, de la que se especializan Agente, Analista y Civil, con el fin de mantener una representación consistente de actores y sujetos implicados sin mezclar responsabilidades operativas con roles de investigación.

La identificación de entidades se realiza mediante identificadores globales estables (por ejemplo, UUID/ULID) asignados en el momento de creación. Se mantiene la inmutabilidad del identificador lógico durante todo el ciclo de vida, con independencia de la tecnología de almacenamiento o de la evolución del esquema. En entidades críticas para trazabilidad (Vestigio, Evento, Análisis) se recomienda adicionalmente un identificador humano-operativo opcional, gestionado como atributo independiente, para facilitar referencia en procedimientos, informes o integraciones.

El versionado se aplica en dos planos. Por un lado, se versiona el esquema del modelo (migraciones controladas y compatibilidad hacia atrás), preservando la interpretabilidad de datos históricos. Por otro lado, se versionan instancias cuando se requiera trazabilidad de modificaciones semánticamente relevantes, evitando sobrescrituras silenciosas de información operativa. En este marco, el cambio de estado se registra como nuevo Evento o como transición derivada de eventos, mientras que correcciones administrativas se registran con metadatos de rectificación y referencia al hecho original.

La nomenclatura se mantiene consistente con el prototipo ontológico: 'Asunto', 'Vestigio', 'Evento', 'Análisis', 'Persona', y sus especializaciones ('VestigioOriginal', 'SubVestigio', así como especializaciones de Evento cuando se incluyan). Se emplean dominios controlados para atributos clasificatorios, con especial atención a 'Vestigio.Tipo', 'Vestigio.Naturaleza', 'Vestigio.Caracter', 'Resultado.GradoCalidad' y 'AfirmacionVinculacion.GradoConfianza', de modo que la explotación posterior no dependa de texto libre no normalizado. La definición de estos dominios se gestiona como catálogo maestro versionado, con equivalencias y deprecaciones explícitas cuando proceda.

### 6.2 Entidades canónicas

En esta subsección se describen las entidades canónicas que estructuran el núcleo del modelo de datos. Se mantienen definiciones alineadas con el Diseño Ontológico, de forma que las entidades preserven su significado operacional y su trazabilidad a lo largo del flujo de trabajo. La materialización en tablas separa conceptos con identidad propia (hechos, objetos trazables y actores) y evita mezclar estado mutable con evidencias auditables. En los puntos en los que existen especializaciones, se adopta un patrón de supertipo-subtipo que permite mantener consistencia semántica y, al mismo tiempo, admitir atributos específicos por rol o por naturaleza del vestigio.

'Asunto' representa la unidad de contexto que delimita el flujo de trabajo y agrupa vestigios y eventos relacionados. Se registran sus metadatos operativos, su estado y la responsabilidad asignada, de forma que

pueda gobernarse el acceso y la explotación posterior sin ambigüedades. La relación entre asunto y vestigio se modela fuera de la tabla de asunto para admitir adscripciones múltiples cuando proceda, pero la entidad asunto conserva atributos propios y ciclo de vida. La estructura se formaliza mediante la tabla [ASUNTO].

Campo	Tipo	Descripción	Restricciones
id	UUID	Identificador único del asunto	PK, NOT NULL
codigo_asunto	VARCHAR(100)	Código humano-operativo del asunto	UNIQUE, NULL
estado	VARCHAR(50)	Estado operativo del asunto	NOT NULL, CHECK IN ('Abierto','EnCurso','Cerrado','Archivado')
clasificacion	VARCHAR(50)	Nivel de clasificación	NOT NULL, CHECK IN ('Normal','Restringido','Confidencial')
responsable_agente_id	UUID	Agente responsable del asunto	FK -> agente(persona_id), NOT NULL
metadata	JSONB	Metadatos del asunto	NULL
version	INTEGER	Versión lógica del registro	NOT NULL, DEFAULT 1
created_at	TIMESTAMP WITH TIME ZONE	Timestamp de creación	NOT NULL, DEFAULT NOW()
updated_at	TIMESTAMP WITH TIME ZONE	Timestamp de última actualización	NOT NULL, DEFAULT NOW()

Tabla [ASUNTO]

'Vestigio' se define como el objeto trazable central. Se registran atributos clasificatorios y descriptivos que permiten distinguir naturaleza y carácter, manteniendo consistencia con dominios controlados, y se asigna responsabilidad operativa conforme al esquema de custodia. El modelo distingue explícitamente entre vestigios incorporados al flujo de trabajo y vestigios derivados, ya que esta distinción determina reglas de trazabilidad y procedencia. En consecuencia, Vestigio se representa como entidad raíz y se extiende en VestigioOriginal y SubVestigio para capturar información específica de entrada o de derivación analítica, sin duplicar campos comunes. Esta estructura se formaliza mediante las tablas [VESTIGIO], [VESTIGIO\_ORIGINAL] y [SUB\_VESTIGIO].

Campo	Tipo	Descripción	Restricciones
-------	------	-------------	---------------

<b>id</b>	UUID	Identificador único del vestigio	PK, NOT NULL
<b>origen_vestigio</b>	VARCHAR(20)	Origen: Original o SubVestigio	NOT NULL, CHECK IN ('Original','SubVestigio')
<b>tipo_vestigio</b>	VARCHAR(20)	Tipo: Físico, digital etc.	NOT NULL
<b>naturaleza</b>	VARCHAR(80)	Naturaleza del vestigio	NOT NULL
<b>caracter</b>	VARCHAR(80)	Carácter: dubitado, indubitado etc.	NOT NULL
<b>descripcion</b>	TEXT	Descripción del vestigio	NULL
<b>responsable_agente_id</b>	UUID	Agente responsable del vestigio	FK -> agente(persona_id), NOT NULL
<b>created_at</b>	TIMESTAMP WITH TIME ZONE	Timestamp de creación	NOT NULL, DEFAULT NOW()
<b>updated_at</b>	TIMESTAMP WITH TIME ZONE	Timestamp de última actualización	NOT NULL, DEFAULT NOW()

Tabla [VESTIGIO]

Campo	Tipo	Descripción	Restricciones
<b>vestigio_id</b>	UUID	Identificador del vestigio	PK, FK -> vestigio(id), NOT NULL
<b>tipo_entrada</b>	VARCHAR(50)	Tipo de entrada (p.ej. ITP, Autopsia)	NOT NULL, CHECK IN ('ITP','Autopsia','Otra')
<b>tipo_entidad</b>	VARCHAR(50)	Tipo de entidad de origen: persona, lugar, vehículo	NOT NULL
<b>evento_entrada_id</b>	UUID	Evento de entrada que lo crea	FK -> evento(id), NOT NULL
<b>metadata</b>	JSONB	Metadatos específicos	NULL

Tabla [VESTIGIO\_ORIGINAL]

Campo	Tipo	Descripción	Restricciones
vestigio_id	UUID	Identificador del vestigio	PK, FK -> vestigio(id), NOT NULL
analisis_id	UUID	Análisis (evento) que lo produce	FK -> analisis(evento_id), NOT NULL
origen_vestigio_id	UUID	Vestigio del que procede (original o subvestigio)	FK -> vestigio(id), NOT NULL
metadata	JSONB	Metadatos específicos	NULL

Tabla [SUB\_VESTIGIO]

'Evento' se define como el hecho auditable que describe acciones sobre vestigios en el marco del flujo de trabajo. Se registran los atributos mínimos para soportar reconstrucción temporal y auditoría (tipo de evento, marcas temporales, ubicación, responsable y huella criptográfica del *payload*). Se evita que la trazabilidad dependa únicamente de estados mutables, y se preserva que el historial se reconstruye a partir de eventos asociados a vestigios. La estructura se formaliza mediante la tabla [EVENTO]. La especialización de Evento en subtipos (por ejemplo, Envío, Almacenamiento, Entrada, Análisis) se implementa con un discriminador o tablas especializadas cuando la variabilidad de atributos lo requiera.

Campo	Tipo	Descripción	Restricciones
id	UUID	Identificador único del evento	PK, NOT NULL
tipo_evento	VARCHAR(50)	Tipo: Recogida, Envío, Almacenamiento, Análisis	NOT NULL, CHECK IN ('Recogida','Envío','Almacenamiento','Análisis')
start_time	TIMESTAMP WITH TIME ZONE	Momento de inicio del evento	NOT NULL
end_time	TIMESTAMP WITH TIME ZONE	Momento de fin del evento	NULL, CHECK (end_time >= start_time)
lugar	VARCHAR(500)	Ubicación donde ocurre el evento	NOT NULL
agente_id	UUID	Agente responsable del evento	FK -> agente(persona_id), NOT NULL
metadata	JSONB	Datos específicos del tipo de evento	NULL
hash_payload	VARCHAR(128)	SHA-512 del payload	NOT NULL

		canonicalizado del evento	
<b>created_at</b>	TIMESTAMP WITH TIME ZONE	Timestamp de creación del registro	NOT NULL, DEFAULT NOW()

Tabla [EVENTO]

'Análisis' se modela como entidad especializada derivada de evento, de modo que hereda temporalidad, ubicación y responsabilidad, y añade estructura propia para gestión técnica (estado del análisis y acreditación obtenida cuando proceda). Esta separación permite tratar el análisis como hecho auditable y, simultáneamente, asociarlo a ejecución por equipo y a protocolos, sin sobrecargar el concepto general de evento. La estructura se formaliza mediante la tabla [ANALISIS], vinculada al evento subyacente por identidad compartida.

Campo	Tipo	Descripción	Restricciones
<b>evento_id</b>	UUID	Identificador del evento de análisis	PK, FK -> evento(id), NOT NULL
<b>estado_analisis</b>	VARCHAR(50)	Estado del análisis	NOT NULL, CHECK IN ('Planificado','EnCurso','Finalizado','Anulado')
<b>acreditacion_id</b>	UUID	Acreditación obtenida (si aplica)	FK -> acreditacion(id), NULL
<b>metadata</b>	JSONB	Metadatos específicos del análisis	NULL
<b>updated_at</b>	TIMESTAMP WITH TIME ZONE	Timestamp de última actualización	NOT NULL, DEFAULT NOW()

Tabla [ANALISIS]

'Persona' constituye el supertipo lógico para representar individuos relevantes en el flujo de trabajo, manteniendo una identificación estable y atributos comunes reutilizables en custodia, ejecución técnica y vinculación. Se registra el tipo de persona mediante un discriminador controlado y se preserva un conjunto mínimo de atributos identificativos conforme a la política aplicable. La especialización en Agente, Analista y Civil permite separar responsabilidades operativas, participación técnica y sujetos vinculados a inteligencia, evitando que un mismo esquema de atributos fuerce interpretaciones incorrectas. La estructura propuesta se formaliza mediante las tablas [PERSONA], [AGENTE], [ANALISTA] y [CIVIL].

Campo	Tipo	Descripción	Restricciones
<b>id</b>	UUID	Identificador único de persona	PK, NOT NULL

<b>tipo_persona</b>	VARCHAR(20)	Tipo: Agente, Analista, Civil	NOT NULL, CHECK IN ('Agente','Analista','Civil')
<b>identificador_externo</b>	VARCHAR(120)	Identificador externo (si existe)	UNIQUE, NULL
<b>nombre</b>	VARCHAR(120)	Nombre	NULL
<b>apellidos</b>	VARCHAR(180)	Apellidos	NULL
<b>documento_ref</b>	VARCHAR(120)	Referencia documental (no necesariamente el número)	NULL
<b>metadata</b>	JSONB	Metadatos adicionales	NULL
<b>version</b>	INTEGER	Versión lógica del registro	NOT NULL, DEFAULT 1
<b>created_at</b>	TIMESTAMP WITH TIME ZONE	Timestamp de creación	NOT NULL, DEFAULT NOW()
<b>updated_at</b>	TIMESTAMP WITH TIME ZONE	Timestamp de última actualización	NOT NULL, DEFAULT NOW()

Tabla [PERSONA]

Campo	Tipo	Descripción	Restricciones
<b>persona_id</b>	UUID	Identificador de persona	PK, FK -> persona(id), NOT NULL
<b>unidad</b>	VARCHAR(120)	Unidad	NOT NULL
<b>id_profesional</b>	VARCHAR(120)	Identificador profesional interno	NULL
<b>activo</b>	BOOLEAN	Indicador de activo	NOT NULL, DEFAULT TRUE
<b>metadata</b>	JSONB	Metadatos del agente	NULL

Tabla [AGENTE]

Campo	Tipo	Descripción	Restricciones
<b>persona_id</b>	UUID	Identificador de persona	PK, FK -> persona(id), NOT NULL
<b>unidad_laboratorio</b>	VARCHAR(120)	Unidad o laboratorio	NOT NULL
<b>activo</b>	BOOLEAN	Indicador de activo	NOT NULL, DEFAULT TRUE
<b>metadata</b>	JSONB	Metadatos del analista	NULL

Tabla [ANALISTA]

Campo	Tipo	Descripción	Restricciones
persona_id	UUID	Identificador de persona	PK, FK -> persona(id), NOT NULL
rol_en_asunto	VARCHAR(50)	Rol dentro del flujo de trabajo del asunto	NULL, CHECK IN ('Investigado','Testigo','Victima','Otro')
metadata	JSONB	Metadatos del civil	NULL

Tabla [CIVIL]

La definición de estas entidades canónicas establece la base estructural del modelo. Cabe señalar que todas las entidades representadas en el Prototipo de Referencia (Sección 5.4.1), incluyendo Protocolo, Acreditación, Reactivo, Equipamiento, Especialidad y sus relaciones asociadas, poseen sus correspondientes tablas en el modelo completo. Sin embargo, por motivos de concisión y para evitar sobrecargar el documento, se ha decidido presentar únicamente las tablas más críticas para el flujo operativo nuclear (Asunto, Vestigio, Evento, Análisis, Persona y sus especializaciones). Las relaciones N:M, las cardinalidades operativas y los mecanismos específicos de trazabilidad y linaje se describen en la subsección siguiente, donde se materializan mediante tablas puente y contratos de integridad asociados.

### 6.3 Relaciones y cardinalidades

En esta subsección se describen las relaciones estructurales entre las entidades canónicas, junto con sus cardinalidades y las implicaciones directas en trazabilidad operativa. Las relaciones se materializan mediante tablas de asociación cuando la cardinalidad es N:M o cuando la relación incorpora semántica propia (por ejemplo, secuenciación). Se mantiene como principio que la trazabilidad se reconstruye a partir de hechos auditables y vínculos explícitos, evitando dependencias en estado mutable o en interpretaciones implícitas.

La relación entre Asunto y Vestigio se define como N:M, de forma que un asunto agrupa múltiples vestigios y un vestigio puede pertenecer a múltiples asuntos cuando exista base procedimental. Esta relación se materializa mediante la tabla [ASUNTO\_VESTIGIO], que actúa como asociación explícita entre [ASUNTO] y [VESTIGIO] y permite incorporar metadatos de pertenencia cuando sea necesario para auditoría. La cardinalidad N:M modelada evita duplicación de vestigios y mantiene independencia entre la identificación del vestigio y su adscripción contextual.

Campo	Tipo	Descripción	Restricciones
asunto_id	UUID	Asunto	PK (compuesta), FK -> asunto(id), NOT NULL
vestigio_id	UUID	Vestigio	PK (compuesta), FK -> vestigio(id), NOT NULL
motivo	VARCHAR(200)	Motivo/justificación operativa	NULL

<b>created_at</b>	TIMESTAMP WITH TIME ZONE	Timestamp de alta	NOT NULL, DEFAULT NOW()
-------------------	--------------------------	-------------------	-------------------------

Tabla [ASUNTO\_VESTIGIO]

La trazabilidad de cadena de custodia se apoya en la relación entre Vestigio y Evento. Un vestigio puede sufrir múltiples eventos a lo largo del flujo de trabajo, y un evento puede afectar a uno o varios vestigios, especialmente en actuaciones operativas donde se gestionan traslados, almacenamientos o intervenciones sobre conjuntos de vestigios. Esta relación N:M se materializa mediante la tabla [EVENTO\_VESTIGIO], que vincula [EVENTO] con [VESTIGIO]. Por ejemplo, un evento de envío puede afectar simultáneamente a 10 vestigios cuando se mueve un lote completo desde una escena al depósito central, generando 10 registros en EVENTO\_VESTIGIO (uno por cada vestigio) con el mismo evento\_id pero diferentes vestigio\_id, manteniendo así la trazabilidad individual y colectiva. Sobre esta base se reconstruye el historial de acciones que han afectado al vestigio, preservando el tipo de evento, la temporalidad, la ubicación y el responsable registrado en la entidad de evento.

Campo	Tipo	Descripción	Restricciones
<b>evento_id</b>	UUID	Evento	PK (compuesta), FK -> evento(id), NOT NULL
<b>vestigio_id</b>	UUID	Vestigio afectado	PK (compuesta), FK -> vestigio(id), NOT NULL
<b>created_at</b>	TIMESTAMP WITH TIME ZONE	Timestamp de creación	NOT NULL, DEFAULT NOW()

Tabla [EVENTO\_VESTIGIO]

Para garantizar una secuenciación determinista de la cadena por vestigio, se introduce una relación adicional con semántica de encadenamiento. La tabla [EVENT\_VESTIGIO\_LINK] registra, para cada par vestigio–evento, el identificador del evento anterior en la secuencia correspondiente a ese vestigio. Esta tabla complementa a [EVENTO\_VESTIGIO] y permite imponer reglas de integridad sobre linealidad, ausencia de bifurcaciones y coherencia temporal cuando se requiera. La secuencia resultante se utiliza para reconstrucción auditada de cadena de custodia y para derivar, como proyección, el estado operativo actual del vestigio en términos de “último evento aplicable” dentro del flujo de trabajo.

Campo	Tipo	Descripción	Restricciones
<b>vestigio_id</b>	UUID	Vestigio	PK (compuesta), FK -> vestigio(id), NOT NULL
<b>evento_id</b>	UUID	Evento actual en la cadena	PK (compuesta), FK -> evento(id), NOT NULL
<b>previous_event_id</b>	UUID	Evento anterior en la cadena (por vestigio)	FK -> evento(id), NULL
<b>created_at</b>	TIMESTAMP WITH TIME ZONE	Timestamp de creación	NOT NULL, DEFAULT NOW()

Tabla [EVENT\_VESTIGIO\_LINK]

La relación de responsabilidad se mantiene de forma consistente en las tres entidades operativas principales: Asunto, Vestigio y Evento. En [ASUNTO] se referencia al Agente responsable mediante su identificador en [AGENTE]; en [VESTIGIO] se registra igualmente el Agente responsable del ítem trazable; y en [EVENTO] se registra el Agente responsable del hecho auditable. Esta coherencia permite evaluar de forma uniforme controles de acceso, auditoría de actuación y trazabilidad de responsabilidades. En este diseño no se introduce una tabla puente específica para responsabilidad (por ejemplo, [AGENTE\_EVENTO]), dado que la responsabilidad se modela como relación 1:N con responsable único: un agente puede ser responsable de múltiples asuntos/vestigios/eventos, y cada asunto/vestigio/evento referencia un único responsable mediante FK. Cuando se requiera registrar participación múltiple (por ejemplo, agentes participantes con rol), se añadirá una relación específica sin sustituir la referencia de responsable único.

En lo relativo al linaje, la distinción entre VestigioOriginal y SubVestigio introduce cardinalidades específicas. Todo registro en [VESTIGIO\_ORIGINAL] referencia un único vestigio raíz en [VESTIGIO] y se asocia a la incorporación inicial del ítem en el flujo de trabajo, manteniendo consistencia con el origen del vestigio. Todo registro en [SUB\_VESTIGIO] referencia un único vestigio raíz en [VESTIGIO] y se vincula al análisis que lo produce, lo que permite identificar de forma inequívoca el origen técnico del derivado. Adicionalmente, la producción de subvestigios se gobierna desde [ANALISIS], que a su vez se define como especialización de [EVENTO], manteniendo la trazabilidad temporal y de responsabilidad del hecho que genera la derivación.

Por otra parte, la relación entre [ANALISTA] y [ANALISIS] se modela mediante [ANALISIS\_ANALISTA], dado que un análisis puede ser realizado por varios analistas y un analista puede participar en múltiples análisis (cardinalidad N:M). La tabla admite además atributos propios de la participación (por ejemplo, rol en el análisis) sin duplicación de registros.

Campo	Tipo	Descripción	Restricciones
analisis_id	UUID	Análisis	PK (compuesta), FK -> analisis(evento_id), NOT NULL
analista_id	UUID	Analista participante	PK (compuesta), FK -> analista(persona_id), NOT NULL
rol_en_analisis	VARCHAR(80)	Rol en el análisis	NULL
created_at	TIMESTAMP WITH TIME ZONE	Timestamp de creación	NOT NULL, DEFAULT NOW()

Tabla [ANALISIS\_ANALISTA]

Las relaciones descritas en esta subsección constituyen el soporte mínimo para trazabilidad de cadena de custodia, reconstrucción temporal de actuaciones y linaje de derivaciones dentro del flujo de trabajo.

#### 6.4 Diagramas entidad-relación (DER)

En esta subsección se presentan los diagramas entidad-relación (DER) del núcleo del modelo de datos. Los diagramas se ofrecen como vista sintética del diseño lógico ya descrito en las subsecciones anteriores, con el objetivo de facilitar la revisión operativa y técnica. Se representan únicamente las entidades y relaciones

necesarias para garantizar trazabilidad en cadena de custodia y trazabilidad de análisis; se omiten catálogos auxiliares, campos de auditoría comunes y metadatos no estructurales para mantener legibilidad. Las relaciones N:M se materializan mediante tablas puente, y las relaciones con semántica de secuencia o procedencia se reflejan explícitamente mediante tablas dedicadas.

#### 6.4.1 DER - Gestión y trazabilidad de la Cadena de Custodia

El DER de cadena de custodia describe el recorrido operativo del vestigio dentro del flujo de trabajo, desde su incorporación y asignación de responsabilidad hasta la secuencia de eventos que constituyen la trazabilidad. El modelo se apoya en EVENTO como hecho auditable y en su vinculación con VESTIGIO mediante EVENTO\_VESTIGIO. Para garantizar reconstrucción determinista de la secuencia por vestigio, la cadena se encadena mediante EVENT\_VESTIGIO\_LINK, donde se registra el evento anterior asociado a un vestigio concreto. La pertenencia de vestigios a asuntos se modela como N:M a través de ASUNTO\_VESTIGIO, permitiendo adscripciones múltiples cuando proceda. La responsabilidad operativa se asigna a AGENTE tanto para ASUNTO como para VESTIGIO y EVENTO, manteniendo un criterio consistente de custodia y auditoría.

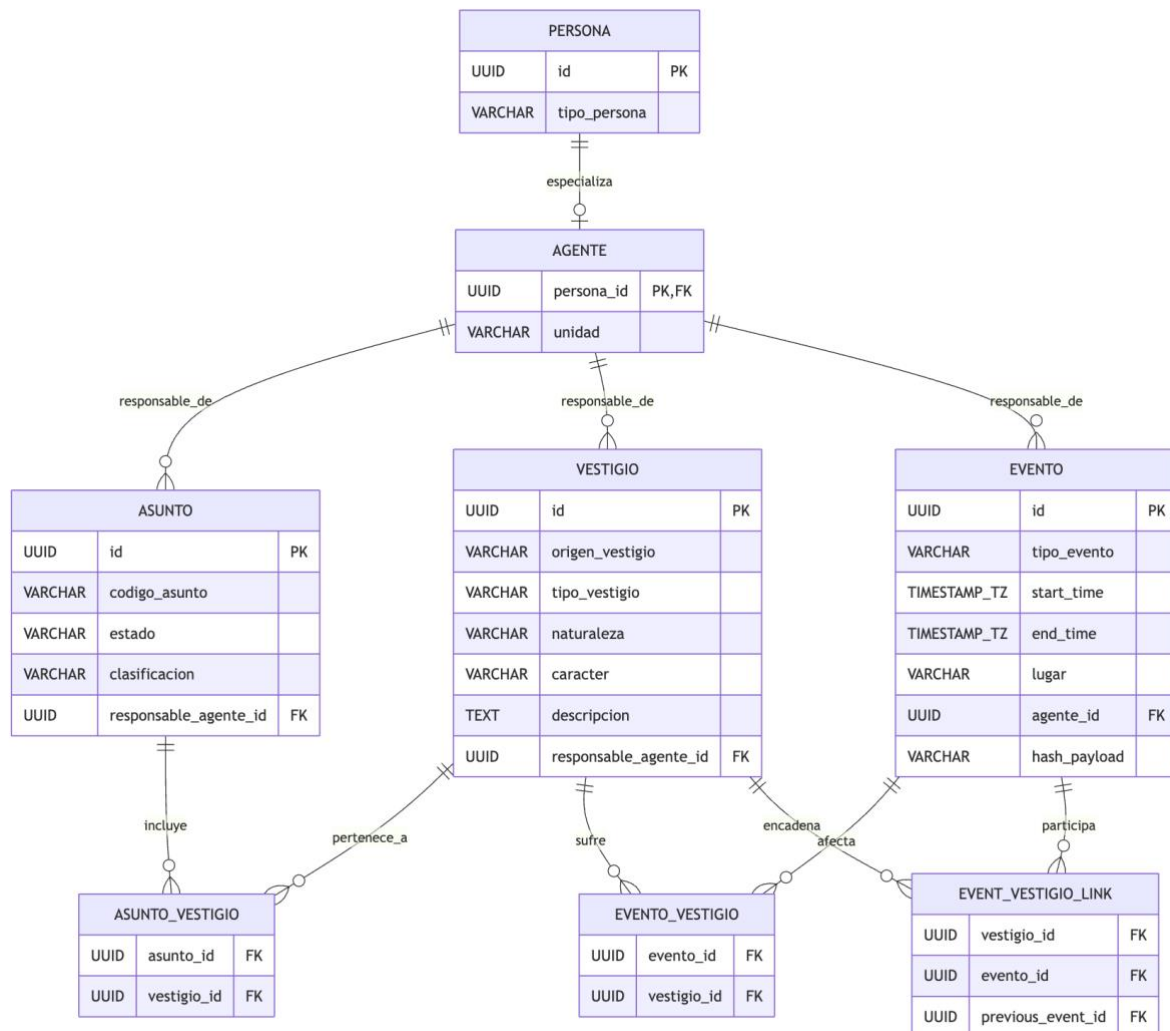


Figura 3 DER Gestión y trazabilidad de la Cadena de Custodia

#### 6.4.2 DER - Gestión y trazabilidad del Análisis de Vestigios

El DER de análisis describe la ejecución técnica sobre vestigios, el personal participante, los protocolos seguidos y los outputs producidos. ANALISIS se modela como subtipo de EVENTO para heredar temporalidad, ubicación y responsabilidad. La relación con ANALISTA se representa como N:M mediante ANALISIS\_ANALISTA, lo que permite reflejar trabajo por equipo y registrar roles operativos dentro del análisis. La relación con PROTOCOLO se representa como N:M mediante ANALISIS\_PROTOCOLO, en coherencia con el requisito de que un análisis pueda seguir varios protocolos.

La producción de resultados se modela con RESULTADO, asociado al análisis productor. La generación de inteligencia se materializa mediante AFIRMACION\_VINCULACION, que vincula un CIVIL y uno o más vestigios, sustentándose en uno o más resultados a través de RESULTADO\_AFIRMACION. El linaje derivado se formaliza con SUB\_VESTIGIO, producido por el análisis, y su procedencia se registra en la propia entidad mediante el atributo origen\_vestigio\_id, referenciado contra VESTIGIO, evitando que la trazabilidad de derivaciones quede implícita y permitiendo que el origen sea tanto un vestigio original como un subvestigio.

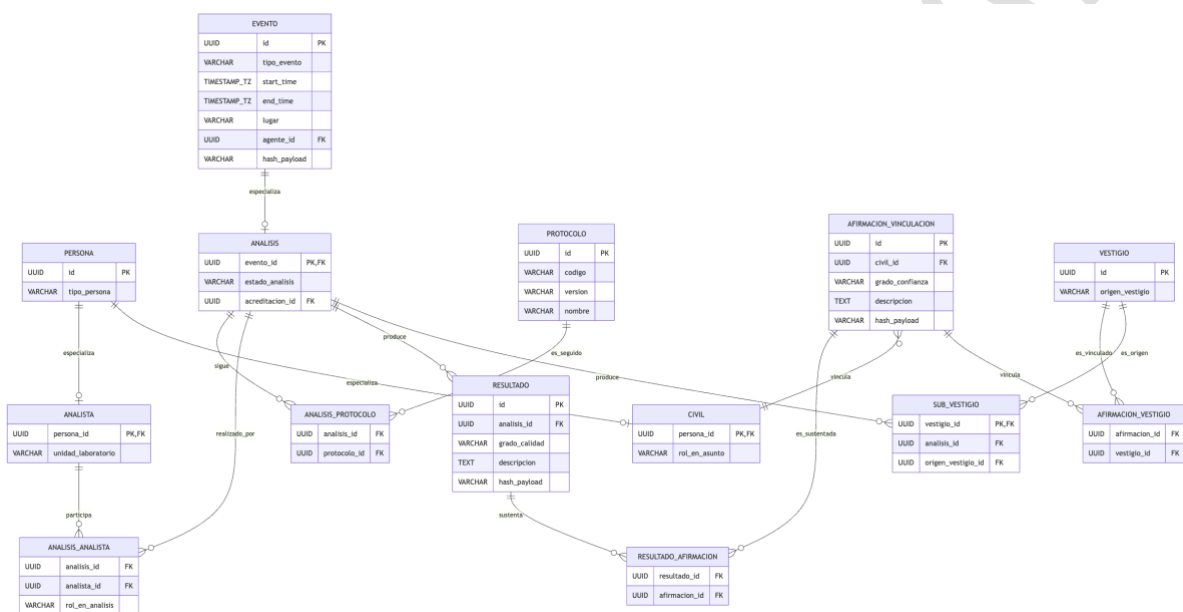


Figura 4 DER Gestión y trazabilidad del Análisis de Vestigios

## 6.5 Reglas de integridad

Las reglas de integridad se formulan como contratos verificables del modelo de datos, de modo que se limite la entrada de estados inconsistentes y se preserve la trazabilidad exigida por el flujo de trabajo forense. Estos contratos se derivan de la estructura ontológica (Vestigio–Evento–Análisis, especializaciones, protocolos, resultados y afirmaciones) y se implementan de forma combinada mediante restricciones en base de datos (PK/FK/UNIQUE/CHECK), validaciones en la capa de dominio y, cuando aplique, validaciones adicionales en repositorios de conocimiento. Se asume que la entidad EVENTO constituye el hecho auditable básico y que la reconstrucción del historial se realiza a partir de relaciones explícitas y secuencias por vestigio. A efectos prácticos, se siguen los siguientes criterios:

- **Integridad referencial (existencia y consistencia de referencias).** Se requiere que todas las claves foráneas apunten a registros existentes y que las especializaciones por identidad compartida se mantengan coherentes. En particular, ANALISIS.evento\_id referencia siempre a EVENTO.id, y el evento asociado debe estar tipado como “Análisis”. De forma análoga, VESTIGIO\_ORIGINAL.vestigio\_id y

SUB\_VESTIGIO.vestigio\_id referencian siempre a VESTIGIO.id. La relación de responsabilidad se mantiene con FK desde ASUNTO.responsable\_agente\_id, VESTIGIO.responsable\_agente\_id y EVENTO.agente\_id hacia AGENTE.persona\_id, garantizando que la responsabilidad se asigna únicamente a agentes válidos.

- Integridad de dominio (valores permitidos y consistencia de tipado). Se aplican dominios controlados para campos clasificatorios y estados operativos. Como mínimo, se restringen mediante CHECK los enumerados de EVENTO.tipo\_evento, ASUNTO.estado, ANALISIS.estado\_analisis, PERSONA.tipo\_persona, y la distinción entre origen del vestigio (original/subvestigio). En el caso de SUB\_VESTIGIO.origen\_vestigio\_id, la referencia se realiza contra VESTIGIO.id para permitir procedencia desde un vestigio original o desde otro subvestigio; adicionalmente, se impone la no autocontención ( $\text{origen\_vestigio\_id} \neq \text{vestigio\_id}$ ) para evitar ciclos triviales. La prevención de ciclos indirectos se controla en capa de dominio, dado que requiere recorrer el grafo de procedencias.
- Integridad temporal (consistencia cronológica y orden). Todo evento debe cumplir  $\text{end\_time} \geq \text{start\_time}$  cuando exista end\_time. En la tabla de encadenamiento EVENT\_VESTIGIO\_LINK se requiere consistencia temporal entre eslabones cuando se registre previous\_event\_id: el start\_time del evento actual no antecede al start\_time del evento previo asociado al mismo vestigio. Esta regla se expresa de forma natural en capa de dominio (o mediante mecanismos avanzados de validación), ya que involucra comparación entre filas de EVENTO vinculadas por previous\_event\_id.
- Contratos de trazabilidad y linaje (reconstrucción determinista). Se exige que todo vínculo de afectación evento–vestigio sea consistente con la secuencia por vestigio. En términos operativos, para cada EVENT\_VESTIGIO\_LINK(vestigio\_id, evento\_id) debe existir el par correspondiente en EVENTO\_VESTIGIO(evento\_id, vestigio\_id). Esta condición puede imponerse con una FK compuesta o mediante validación transaccional en la capa de persistencia. Adicionalmente, se recomienda restringir bifurcaciones: para un vestigio dado, un evento\_id no debe tener más de un predecesor registrado, y un mismo previous\_event\_id no debe apuntar a múltiples “siguientes” cuando se exija cadena lineal. Estas restricciones se implementan con índices UNIQUE adecuados sobre combinaciones de campos en EVENT\_VESTIGIO\_LINK, en función del criterio operativo acordado (cadena estrictamente lineal o cadena con ramificaciones explícitas).
- Contratos específicos de análisis (equipo y procedimientos). Dado que un análisis puede ser realizado por varios analistas, se materializa la relación N:M mediante ANALISIS\_ANALISTA, y se impone unicidad de pares (analisis\_id, analista\_id). Si se requiere que todo análisis tenga al menos un analista asignado, se valida en capa de dominio en el momento de cierre o transición a estados finales. De forma equivalente, si todo análisis debe seguir al menos un protocolo, se valida sobre la relación N:M ANALISIS\_PROTOCOLO en los hitos de transición del análisis. Estas reglas no se expresan de forma robusta con CHECK a nivel de fila, ya que dependen del conteo de filas relacionadas, por lo que se implementan como invariantes transaccionales en la lógica de dominio.
- Coherencia entre resultados, afirmaciones y vestigios vinculados. Un RESULTADO debe referenciar siempre a un ANALISIS existente. Una AFIRMACION\_VINCULACION referencia a un CIVIL existente y se vincula a uno o más vestigios mediante su tabla de asociación; además, su sustentación se registra mediante RESULTADO\_AFIRMACION, que permite trazar qué resultados sustentan qué afirmaciones. Si se exige que toda afirmación tenga al menos un resultado sustentante y al menos un vestigio vinculado, se valida en capa de dominio (p. ej., antes de permitir su estado “emitida” o equivalente). La calidad (grado\_calidad) y confianza (grado\_confianza) se restringen a dominios controlados para facilitar explotación y evitar texto libre no normalizado.
- Integridad criptográfica y auditabilidad mínima. En EVENTO.hash\_payload se requiere no nulidad y longitud fija según el algoritmo acordado; el payload se canonicaliza antes de calcular el hash para evitar inconsistencias por serialización. Para preservar la naturaleza auditable del evento, se recomienda tratar los eventos como registros inmutables: las correcciones operativas se reflejan mediante nuevos

eventos y no mediante actualización in situ, reservando las actualizaciones directas para metadatos administrativos no probatorios cuando proceda.

Con este enfoque, los contratos no se limitan a “restricciones de tabla”, sino que se convierten en reglas operativas verificables en transacciones: se validan en el punto en el que el flujo de trabajo produce efectos (alta de vestigio, incorporación a asunto, registro de evento, cierre de análisis, emisión de afirmación), reduciendo el riesgo de datos parcialmente válidos que impidan auditoría o reconstrucción posterior.

## 6.6 Principios seguidos

El modelo propuesto se define como un núcleo canónico orientado a trazabilidad, auditoría y explotación, y se mantiene independiente de una tecnología concreta de persistencia. Esta neutralidad se fundamenta en separar la semántica del dominio (entidades, relaciones y contratos) de su materialización física (tablas, grafos, índices, proyecciones o flujos de eventos). En consecuencia, se conserva una representación estable de identificadores, relaciones y linaje, y se habilita que distintas tecnologías consuman el mismo significado sin reinterpretaciones locales. Algunos de los principios tenidos en cuenta son:

- **Agnosticismo tecnológico y proyecciones especializadas.** Se preserva una “fuente de verdad” coherente (entidades canónicas y eventos auditables) y se permite construir proyecciones para necesidades específicas: consultas transaccionales, exploración relacional, consultas de grafo, búsqueda semántica o auditoría inmutable. La solución mantiene los identificadores y las relaciones como invariantes, de forma que un mismo Vestigio.id o Evento.id se reutiliza en proyecciones sin duplicidad semántica. Este principio reduce fricción en integraciones y facilita evolución del almacenamiento sin reescritura conceptual del dominio.
- **Eventos como unidad auditable y estado como proyección.** Se registra el flujo de trabajo mediante EVENTO y sus relaciones con vestigios, de forma que el historial sea reconstruible y verificable. El “estado actual” (por ejemplo, el último evento aplicable o el responsable vigente) se interpreta como una proyección derivada del historial, evitando que el sistema dependa de estados mutables como fuente única. Este criterio simplifica auditorías y permite reproducibilidad ante revisiones.
- **Linaje explícito y derivaciones gobernadas.** Se formaliza la distinción entre vestigio original y subvestigio y se hace explícita la procedencia mediante SUB\_VESTIGIO.origen\_vestigio\_id. Con ello se evita que el linaje quede implícito en texto o en metadatos dispersos, y se habilita navegación determinista de derivaciones (hacia atrás y hacia adelante) para trazabilidad técnica y explotación.
- **Consistencia por contratos y validación en hitos.** Se asume que no todas las restricciones se expresan de forma suficiente en el esquema relacional. Por ello, se aplican contratos en el momento operativo relevante: registro de evento, cierre de análisis, emisión de afirmación, etc. Este enfoque reduce estados intermedios incoherentes y alinea el modelo con ejecución real del flujo de trabajo.
- **Evolución controlada y compatibilidad.** Se versionan dominios controlados y, cuando proceda, instancias lógicas. Se evitan cambios destructivos no trazables; las migraciones de esquema se gestionan de forma controlada, preservando interpretabilidad de datos históricos. Este principio es especialmente relevante en entidades auditables y en relaciones de encadenamiento, donde la pérdida de información compromete reconstrucción de historial.
- **Minimización y compartimentación de datos personales.** Se centraliza la identidad en PERSONA y se separan roles en tablas especializadas, de modo que se reduzcan duplicidades y se facilite el gobierno de acceso y minimización. Este criterio habilita segmentación por permisos (por ejemplo, visibilidad limitada de atributos identificativos) sin romper trazabilidad, dado que las referencias se mantienen por identificador.
- **Interoperabilidad semántica.** La estructura se mantiene alineada con el diseño ontológico de referencia, lo que permite trazar correspondencias hacia representaciones semánticas (por ejemplo, OWL/SHACL)

cuando se requiera interoperabilidad, validación adicional o explotación basada en conocimiento. Esta alineación reduce ambigüedad terminológica y facilita la integración posterior con repositorios de conocimiento y servicios de consulta avanzados.

CONFIDENCIAL

## 7 DESTINO Y ALMACENAMIENTO DE LOS DATOS

### 7.1 Objetivo y alcance de la persistencia

En este documento se entiende por destino de datos el conjunto de repositorios donde se materializan, con propósitos diferenciados, los registros necesarios para operar el flujo de trabajo forense, reconstruir la trazabilidad y sostener evidencia verificable. Se separa explícitamente la fuente de verdad (registro canónico transaccional) de las proyecciones (repositorios derivados optimizados para consulta y explotación), y de la evidencia verificable (registro inmutable orientado a valor probatorio). Esta separación evita que la auditoría dependa de estados mutables y permite que la reconstrucción histórica se apoye en hechos registrados y correlacionables.

La fuente de verdad se ubica en almacenamiento relacional con garantías ACID para el modelo canónico (por ejemplo, entidades y relaciones operativas como VESTIGIO, EVENTO, ANALISIS, RESULTADO y AFIRMACION\_VINCULACION), de modo que las operaciones de alta y vinculación se registran con integridad referencial y restricciones de consistencia. Sobre esta base, se registran también las huellas criptográficas (hash\_payload) asociadas a eventos y artefactos críticos, preservando la capacidad de verificación posterior sin condicionar el rendimiento de consulta.

Las proyecciones se definen como réplicas derivadas para optimización de consulta y servicios avanzados. En particular, la base de datos de grafos se utiliza para materializar relaciones complejas y recorridos (por ejemplo, caminos entre vestigios, eventos, análisis, resultados, afirmaciones y personas) alimentándose de eventos y/o cambios confirmados en el canónico. La base de datos vectorial se emplea como índice semántico de campos textuales relevantes (descripciones, contenido técnico, interpretaciones) para búsquedas por similitud, manteniendo la correlación por identificadores canónicos. Asimismo, el almacenamiento de objetos se reserva para evidencias binarias (imágenes, vídeos, PDFs firmados, datos brutos instrumentales), guardando en el canónico los metadatos, referencias y hashes de integridad necesarios para control y verificación.

La evidencia verificable se articula como un plano separado, denominado inmutable judicial, donde se preservan pruebas criptográficas y anclajes asociados a eventos y artefactos con valor probatorio. En el diseño se establece que determinados payloads canonicalizados (por ejemplo, eventos de cadena de custodia, análisis, resultados y afirmaciones) se hashean (SHA-512) y se registran de forma inmutable, incorporando además mecanismos de anclaje externo para refuerzo de no repudio. Esta capa no sustituye al canónico, sino que lo complementa: el canónico soporta operación y consulta; el inmutable judicial soporta verificación y evidencia.

De forma operativa, el alcance de la persistencia se fija separando cinco categorías, cada una con su destino primario y su función dominante: (i) datos transaccionales (canónico relacional), (ii) evidencias binarias (objetos + metadato y hash en canónico), (iii) eventos/auditoría (event store + tabla de eventos y encadenamientos), (iv) conocimiento/relaciones (grafo como proyección), y (v) capacidades de búsqueda (vectorial como proyección).

### 7.2 Principios de diseño para almacenamiento y custodia

El diseño de almacenamiento se alinea con el modelo de datos mediante un principio rector: la trazabilidad se reconstruye a partir de eventos. Los eventos constituyen el mecanismo primario de auditoría, porque encapsulan quién ejecuta una acción, cuándo ocurre, dónde ocurre y sobre qué entidad impacta, permitiendo además encadenamiento lógico para secuenciación verificable (por ejemplo, previous\_event\_id en eventos de custodia). Esta aproximación soporta *event sourcing* (patrón arquitectónico donde el estado del sistema se reconstruye a partir de una secuencia inmutable de eventos en lugar de depender de estados mutables almacenados directamente) y facilita proyecciones consistentes hacia repositorios secundarios, sin introducir dependencia funcional de estados calculados.

La inmutabilidad probatoria se formula como un compromiso criptográfico separado de la explotación operativa. Se registran hashes (`hash_payload`) sobre payloads canonicalizados de los elementos probatoriamente relevantes, y se preservan anclajes y referencias de verificación en un plano inmutable judicial, con correlación por identificadores canónicos. De este modo, la plataforma mantiene la capacidad de verificación de integridad y no repudio sin imponer que todos los repositorios operen como *ledgers*, y sin degradar los patrones de acceso necesarios para consulta y gestión diaria.

La persistencia se diseña como polígota, asignando cada repositorio al patrón de acceso que optimiza. El canónico relacional absorbe transacciones y restricciones de integridad; la base de datos de eventos (*Event Store*) soporta flujos inmutables y desacoplo entre dominios mediante suscripción; el grafo optimiza recorridos y consultas relacionales profundas; el vectorial soporta recuperación semántica; y el almacenamiento de objetos absorbe binarios a gran escala con versionado y políticas de retención. La sincronización entre repositorios se ejecuta mediante consumo de eventos y controles de idempotencia, y se refuerza con patrones como Outbox para garantizar emisión tras *commit* cuando proceda.

Finalmente, se aplica una separación estricta dato–metadato–prueba. Los binarios se conservan en objetos; el metadato operativo y las relaciones (incluida la referencia al objeto y su *hash*) se mantienen en el canónico; y las pruebas criptográficas (hashes, anclajes, identificadores de transacción) se registran en el inmutable judicial y en tablas de correlación diseñadas para auditoría. Esta separación mantiene claridad técnica sobre qué se consulta para operar, qué se verifica para probar, y qué se preserva como evidencia binaria, evitando ambigüedades en la cadena de custodia y facilitando controles de consistencia entre planos.

### 7.3 Matriz de asignación de destino para los datos

En esta subsección se fija, para cada categoría de datos, un destino primario, una función dominante y una correlación mínima basada en identificadores canónicos. De esta manera, la Tabla 2 recoge la Matriz de asignación de destino para los datos, estableciendo la separación entre datos transaccionales canónicos (núcleo relacional), evidencias binarias (almacenamiento de objetos), eventos/auditoría (*event store* y tablas de apoyo), conocimiento/relaciones (grafo como proyección) y capacidades de búsqueda (vectorial como proyección). La categoría de “evidencia verificable” se trata de forma diferenciada: no se utiliza como repositorio operativo, sino como plano probatorio, donde se registran compromisos criptográficos (hashes sobre payloads canonicalizados) y, cuando proceda, pruebas de anclaje. Esta separación preserva el rendimiento y la flexibilidad del plano operativo, sin diluir el requisito de no repudio y verificación independiente en cadena de custodia.

Operativamente, la correlación mínima indicada en la matriz (UUID canónicos y claves compuestas de encadenamiento) se usa como conexión entre repositorios. Por ejemplo, el almacenamiento de objetos conserva binarios con su `object_key/object_id`, mientras que el canónico mantiene la referencia y el hash; el plano inmutable judicial conserva el compromiso criptográfico y, en su caso, el anclaje, permitiendo verificar a posteriori que la evidencia consultada corresponde a la registrada. En el mismo sentido, el grafo y el vectorial se consideran proyecciones reconstruibles, por lo que su consistencia se gobierna por idempotencia y reindexación controlada, no por autoridad de escritura.

Categoría	Qué incluye	Destino	Función dominante	Correlación mínima
Datos canónicos	ASUNTO, VESTIGIO, PERSONA, EVENTO, ANALISIS y relaciones operativas	Relacional	Integridad, transacciones, consistencia	UUID canónicos ( <code>asunto_id</code> , <code>vestigio_id</code> , <code>evento_id</code> , <code>persona_id</code> )

<b>Evidencias binarias</b>	Imágenes, vídeo, PDFs, ficheros instrumentales, informes firmados	Objetos	Persistencia de binarios, versionado/re tención	Referencia a objeto + hash en canónico (p.ej. hash_payload)
<b>Eventos y auditoría</b>	Eventos operativos y de custodia; secuencias por vestigio	Event Store + Relacional (EVENTO)	Histórico auditable, proyecciones	evento_id, vestigio_id, previous_event_id
<b>Conocimiento y relaciones</b>	Relaciones complejas y recorridos (vestigio→evento→análisis→resultado→afirmación→persona)	Grafos (proyección)	Consultas por recorridos y exploración	Reutilización de UUID canónicos
<b>Capacidades de búsqueda</b>	Búsqueda semántica sobre textos operativos/técnicos	Vectorial (proyección)	Recuperación por similitud	entity_id, entity_type, embedding_id
<b>Evidencia verificable</b>	Compromisos criptográficos y anclajes	Inmutable judicial	Verificación, no repudio	event_id + hash_payload + prueba/anclaje

Tabla 2 Matriz de asignación de destino

#### 7.4 Rutas y flujos de sincronización

La Tabla 3 describe cómo se propagan hechos y proyecciones entre destinos, manteniendo un principio rector: los repositorios secundarios no se consideran fuente de verdad, sino materializaciones derivadas que se reconstruyen desde el canónico y desde el plano de eventos cuando aplica. Esta tabla se puede interpretar como un contrato de integración: para cada ruta se define el mecanismo (proyección derivada, indexación, anclaje, consumo de eventos), la unidad de publicación (identificadores y relaciones), la garantía mínima (idempotencia, orden por *stream*, reindexación controlada) y el uso principal.

La proyección a grafo se alimenta desde EVENTO y sus vínculos (incluyendo, cuando se usa, el encadenamiento por vestigio). La idempotencia por evento\_id y por claves de relación permite reprocesar sin duplicidades, lo que resulta crítico ante reintentos, replays o reconstrucciones completas. La proyección vectorial se alimenta desde textos del canónico (y, cuando proceda, desde metadatos de resultados e informes), manteniendo entity\_id y entity\_type como claves de reindexación; de este modo, una regeneración de embeddings no altera la identidad de la entidad, sino su representación en búsqueda.

La ruta hacia el inmutable judicial se trata como un flujo separado del resto: no se replica “dato operativo” sino prueba verificable (*hash* + referencia + metadatos de anclaje). En términos de implementación, la secuencia recomendada mantiene primero la persistencia en canónico y el registro del evento, y después el cálculo y registro del compromiso criptográfico, evitando que el plano probatorio se convierta en un cuello de botella del plano transaccional. Cuando se usan patrones de emisión robusta (por ejemplo, Outbox), se preserva la correspondencia entre commit transaccional y publicación de eventos, reduciendo el riesgo de divergencias entre destinos.

Todo este proceso se plasma en un diagrama de sincronización representado en la Figura 5.

Origen	Destino	Mecanismo	Unidad publicación	Uso principal
--------	---------	-----------	--------------------	---------------

<b>Relacional (EVENTO + vínculos)</b>	Grafos	Proyección derivada	evento_id y relaciones	Consultas por recorrido
<b>Relacional (textos)</b>	Vectorial	Indexación	entity_id + contenido	Búsqueda semántica
<b>Objetos (binarios) + canónico (metadatos)</b>	Inmutable judicial	Anclaje	hash + referencia	Cadena de custodia probatoria
<b>Event Store</b>	Relacional, grafo, vectorial	Consumo de eventos	Evento (append-only)	Reconstrucción y auditoría

Tabla 3 Rutas y flujos de sincronización

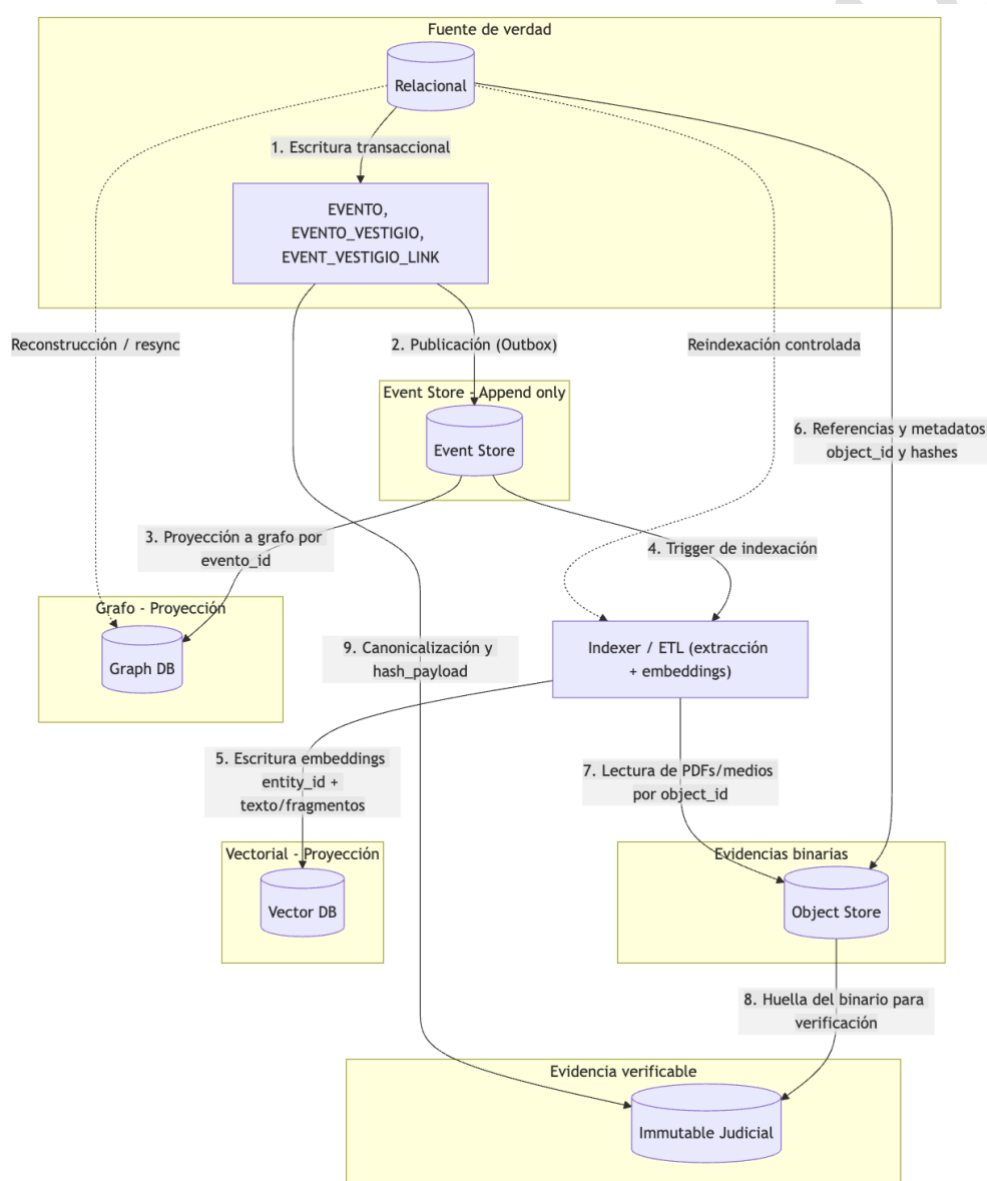


Figura 5 Diagrama de sincronización de flujos de datos

## 8 MANTENIMIENTO DE LA CADENA DE CUSTODIA

### 8.1 Glosario de términos

Con el fin de mantener consistencia terminológica a lo largo del documento, se establecen las siguientes convenciones:

- **Payload canonicalizado:** Representación normalizada de un conjunto de datos en la que se estandarizan el orden de campos, formatos de fecha y codificación, permitiendo calcular huellas criptográficas idénticas para datos semánticamente equivalentes.
- **Hash payload / Huella criptográfica:** En el texto narrativo se utiliza preferentemente "huella criptográfica" para referirse al resultado de aplicar una función hash (SHA-512) sobre el *payload* canonicalizado. En tablas, código y esquemas técnicos se emplea el término ``hash_payload`` como nombre de campo.
- **Event sourcing:** Patrón arquitectónico en el que el estado del sistema se reconstruye a partir de una secuencia inmutable de eventos, en lugar de almacenar directamente estados mutables. Permite auditoría completa y reproducibilidad del historial.
- **Inmutable judicial:** Plano de almacenamiento separado donde se preservan pruebas criptográficas (hashes, anclajes, identificadores de transacción) y referencias de verificación para artefactos con valor probatorio, complementando el almacenamiento canónico operativo.

### 8.2 Estrategia

El mantenimiento de la cadena de custodia se formaliza como la capacidad de (i) reconstruir secuencias operativas a partir de hechos auditables y (ii) verificar, de forma independiente, la integridad de los artefactos probatoriamente relevantes. En este marco, la Tabla 4 fija qué elementos se consideran anclables, cuál es su identificador de correlación, qué huella se calcula (por ejemplo, `hash_payload`), en qué momento se registra y cuál es su objetivo probatorio. Esta estructura evita que la cadena de custodia dependa de estados derivados no verificables y permite auditorías repetibles sobre la misma base factual.

A partir del prototipo de ontología propuesto, se deduce que la unidad probatoria primaria en el plano operativo es el EVENTO, junto con su posible encadenamiento por vestigio (por ejemplo, mediante `EVENT_VESTIGIO_LINK`). Para cada evento relevante se canonicaliza el *payload* antes de calcular el hash, de manera que la huella no dependa del orden de campos, formatos de fecha o serializaciones equivalentes. El hash resultante se registra en el canónico como parte del propio evento y se correlaciona con el plano inmutable judicial, donde se conserva el compromiso criptográfico y, cuando se aplica, su anclaje. Con ello se separa claramente "lo que se consulta para operar" (canónico y proyecciones) de "lo que se verifica para probar" (inmutable judicial), manteniendo trazabilidad y verificabilidad sin forzar que todos los repositorios operen como libro mayor.

En términos de verificación, se aplica un procedimiento repetible: (1) se recupera el registro operativo (EVENTO, RESULTADO, AFIRMACION\_VINCULACION o referencia de objeto), (2) se reconstruye el *payload* canonicalizado conforme a la misma especificación de canonicalización, (3) se recalcula la huella y se compara con `hash_payload`, y (4) se valida la prueba/anclaje registrada en el inmutable judicial para confirmar no repudio. Para binarios críticos, el hash del fichero se calcula sobre el contenido y se conserva tanto en el canónico (asociado a la referencia de objeto) como en el plano inmutable; de este modo se verifica que el fichero servido por almacenamiento de objetos coincide con el registrado como evidencia.

Finalmente, cuando se requiere rectificación operativa (por ejemplo, correcciones administrativas), la integridad probatoria se mantiene evitando la sustitución silenciosa: el hecho original permanece y la rectificación se registra como nuevo hecho vinculado, conservando referencias cruzadas y huellas correspondientes. Este criterio alinea el mantenimiento de cadena con un modelo de auditoría por eventos y con la separación dato–metadato–prueba descrita en la Sección 7, preservando coherencia técnica entre persistencia, sincronización y verificabilidad.

Artefacto/registro	Identificador	Huella	Momento de anclaje	Objetivo probatorio
<b>Evento de custodia (EVENTO)</b>	evento_id	hash_payload (SHA-512)	En el alta/confirmación del evento	Integridad y no repudio del hecho
<b>Encadenamiento por vestigio (EVENT_VESTIGIO_LINK)</b>	(vestigio_id,evento_id)	Hash del enlace o del evento	Al registrar el enlace	Verificación de secuencia
<b>Binario crítico (objeto)</b>	object_key / object_id	sha256/sha512 del binario	En el alta del binario	Integridad del fichero
<b>Resultado (RESULTADO)</b>	resultado_id	hash_payload	En la finalización del resultado	Integridad del output
<b>Afirmación de vinculación</b>	afirmacion_id	hash_payload	En la emisión/confirmación	Integridad de la afirmación

Tabla 4 Resumen estrategia probatoria

## 9 INTEROPERABILIDAD

La plataforma se diseñará para digitalizar y automatizar los intercambios de información científico-forense, internos y con sistemas externos, empleando formatos estandarizados y respetando plazos de respuesta y condiciones legales del procedimiento. La interoperabilidad se regirá por el Documento de Arquitectura de Interoperabilidad (DAI) y un Plan de Pruebas de Interoperabilidad (PPI).

### 9.1 Interoperabilidad semántica y formatos de intercambio

La interoperabilidad semántica se define como la capacidad de intercambiar información preservando su significado operacional, de forma que diferentes sistemas puedan interpretar los mismos datos con la misma lógica de dominio. En el contexto de THOT, esta interoperabilidad se apoya en dos decisiones ya adoptadas en el documento: (i) un modelo canónico con identificadores estables y relaciones explícitas, y (ii) una trazabilidad reconstruible por eventos (EVENTO y vínculos) como soporte primario de auditoría y linaje.

A efectos prácticos, “interoperar semánticamente” no se limita a exportar JSON o CSV; implica publicar entidades (p. ej., VESTIGIO, EVENTO, ANALISIS) como recursos identificables y enlazables, con vocabularios y tipos controlados, y con un mecanismo verificable para comprobar integridad cuando la información se utiliza fuera de la plataforma (por ejemplo, mediante hash\_payload y anclajes en el plano inmutable judicial, cuando proceda).

Para materializar la interoperabilidad semántica se utiliza JSON-LD como formato de intercambio. JSON-LD permite mantener la ergonomía de JSON (fácil de consumir por APIs y servicios) y, a la vez, añadir semántica mediante un contexto (@context) que mapea claves JSON a identificadores de vocabulario (IRIs). Con ello se obtiene:

- Identificación global de entidades mediante @id (alineable con los UUID/ULID canónicos ya definidos).
- Tipado explícito mediante @type (por ejemplo, thot:Vestigio, thot:Evento, thot:Analisis).
- Enlaces entre entidades sin duplicación (por ejemplo, un EVENTO referencia a un VESTIGIO por su @id, y no por copia de atributos).
- Extensibilidad controlada: un sistema externo puede ignorar propiedades no reconocidas sin perder el núcleo semántico.

Se recomienda publicar un contexto versionado (por ejemplo, thot-context/v1) y mantener una política de compatibilidad hacia atrás similar a la ya adoptada para el esquema del modelo de datos.

Además, para evitar ambigüedades, se estructura el intercambio con un “sobre” que separa: (i) el recurso principal, (ii) sus referencias mínimas (ids enlazables), y (iii) los metadatos de auditoría/verificación cuando aplique. Esta separación es consistente con el principio dato–metadato–prueba y con la existencia del plano inmutable judicial como evidencia verificable.

Un ejemplo simplificado de VESTIGIO en JSON-LD, enlazado a un ASUNTO y con referencia a su responsable:

```
{
  "@context": {
    "thot": "https://thot.example/ns#",
    "id": "@id",
    "type": "@type",
    "Asunto": "thot:Asunto",
```

```

    "Vestigio": "thot:Vestigio",
    "perteneceA": {"@id": "thot:pertenece_a", "@type": "@id"},
    "responsable": {"@id": "thot:responsable_de", "@type": "@id"},
    "tipo": "thot:Tipo",
    "naturaleza": "thot:Naturaleza",
    "caracter": "thot:Caracter",
    "descripcion": "thot:Descripcion"
  },
  "id": "urn:uuid:2f1a7e2c-0d4a-4c9e-9a70-5b6f3d1b5c8a",
  "type": "Vestigio",
  "tipo": "Biologico",
  "naturaleza": "Fisico",
  "caracter": "Muestra",
  "descripcion": "Hisopo recogido en laboratorio móvil",
  "perteneceA": [
    "urn:uuid:8a0d2d6e-9b7a-4d07-9a8a-6b91c2c4d122"
  ],
  "responsable": "urn:uuid:6b0e0f3d-4d21-4a2a-8b16-91a8a9c9d0ee"
}

```

## 9.2 Interoperabilidad entre repositorios de la plataforma THOT

La plataforma ya separa destinos de persistencia por categoría (canónico relacional, objetos, event store, grafo, vectorial e inmutable judicial). En este contexto, JSON-LD actúa como:

- Formato de salida común desde el canónico relacional (fuente de verdad) para APIs y conectores.
- Formato de publicación para proyecciones derivadas, cuando interese exponer recorridos del grafo o resultados de búsqueda vectorial sin acoplar al motor subyacente.
- Formato de evidencia exportable para auditoría y verificación, en combinación con hash\_payload y el registro de prueba/anclaje en el plano inmutable judicial cuando el flujo de trabajo lo exija.

En particular, el componente EVENTO y sus enlaces son adecuados para intercambio porque encapsulan quién, cuándo, dónde y sobre qué entidad impacta una acción, permitiendo reconstrucción y validación de trazabilidad fuera de THOT si se conserva el encadenamiento y la integridad del *payload*.

Este es un ejemplo de EVENTO (cadena de custodia) que afecta a un vestigio y conserva huella criptográfica:

```

{
  "@context": {
    "thot": "https://thot.example/ns#",
    "id": "@id",
    "type": "@type",
    "Evento": "thot:Evento",

```

```

"afectaA": {"@id": "thot:sufre", "@type": "@id"},
"responsable": {"@id": "thot:responsable_de", "@type": "@id"},
"fechaInicio": "thot:FechaInicio",
"fechaFin": "thot:FechaFin",
"lugar": "thot:Lugar",
"previousEvent": {"@id": "thot:previous_event_id", "@type": "@id"},
"hashPayload": "thot:hash_payload"
},
"id": "urn:uuid:9f3f5aa0-2a1e-4f2c-b4c9-2b6a3f3a2e10",
"type": "Evento",
"fechaInicio": "2026-01-24T10:12:00Z",
"fechaFin": "2026-01-24T10:20:00Z",
"lugar": "Depósito central",
"responsable": "urn:uuid:6b0e0f3d-4d21-4a2a-8b16-91a8a9c9d0ee",
"afectaA": "urn:uuid:2f1a7e2c-0d4a-4c9e-9a70-5b6f3d1b5c8a",
"previousEvent": "urn:uuid:1b1e1c3d-9a32-4b10-a1b2-0c9a9b2a1d77",
"hashPayload": "sha512:7c3a...e91f"
}

```

### 9.3 Interoperabilidad entre bases de datos de PN

La interoperabilidad interna se basará en la integración estructurada entre la plataforma THOT y los sistemas corporativos de la Policía Nacional, así como con los grandes repositorios forenses ya existentes, garantizando en todo momento la interoperabilidad dentro del ecosistema policial sin duplicidades ni reprocesos manuales, y alineándose con los requisitos del pliego (INTEROP) y con los flujos operativos descritos en las jornadas; en este marco se contemplan, al menos, los siguientes sistemas:

- **Sistemas de identidad e historial policial:** PERSONAS como núcleo de filiación, reseñas e historial.
- **Sistemas biométricos y genéticos:** ABIS corporativo (huellas, palmares, facial) y CODIS nacional (perfiles genéticos).
- **Sistemas balísticos y forenses especializados:** IBIN nacional y otros sistemas de balística, así como los módulos de disciplinas forenses que se articulan alrededor de BINCIPO.
- **Sistemas de gestión operativa y administrativa:** BINCIPO como eje de gestión de asuntos, vestigios, actividades y envíos; otros sistemas corporativos relevantes (p.ej. ADEXTTRA u otros vinculados a extranjería y gestión operativa).

THOT se apoya en este ecosistema para optimizar el ciclo de vida del dato forense y operativo. El diseño de datos se organiza alrededor de:

- Un **identificador único de persona** (ordinal en PERSONAS), reutilizado en los diferentes procedimientos (reseñas de detenidos, extranjería, asilo, etc.), evitando duplicidades de identidad.
- Un **identificador único de Asunto**, que articula la información gestionada en BINCIPO (entradas, diligencias, filiaciones, direcciones, vestigios, actividades, envíos, adjuntos) y que THOT reutiliza como eje de correlación.

- **Identificadores únicos de vestigio, escena y actividad analítica**, que permiten vincular de forma consistente los resultados de ABIS, CODIS, IBIN y otros sistemas especializados con las personas y asuntos correspondientes.

En la práctica, esto se traduce en la automatización de los procesos de intercambio de información científico-forense entre bases de datos internas:

- Los **datos capturados en reseñas, inspecciones técnico-policiales o procedimientos de identificación** (p.ej. asilo o entradas irregulares) se registran una única vez y se reutilizan en PERSONAS, BINCIPOL y los repositorios biométricos/genéticos.
- Los **resultados de cotejos biométricos o genéticos** (*matches* en ABIS, hits en CODIS, correlaciones balísticas en IBIN) se integran automáticamente en el Asunto correspondiente, sin necesidad de reprocesos manuales ni cargas duplicadas.
- La información forense relevante se **expone hacia el sistema de inteligencia policial (GATI/Investiga)** como eventos estructurados (vestigios con identidad, vestigios anónimos, coincidencias nacionales o internacionales), evitando el doble registro manual que existe actualmente entre BINCIPOL y los sistemas de inteligencia.

De este modo, la interoperabilidad interna no se limita a la conectividad técnica, sino que establece un **modelo de datos compartido** (persona–asunto–vestigio–resultado) que se aplica de forma homogénea en los diferentes casos de uso (reseña de detenidos, extranjería/asilo, ITP, ciclos de vida de vestigios, etc.), asegurando que todos los sistemas “hablan” del mismo caso y de las mismas entidades.

#### 9.4 Interoperabilidad entre bases de datos externas

Además de la integración interna, THOT debe operar en un contexto de criminalidad transnacional y cooperación interinstitucional, lo que exige una interoperabilidad avanzada con sistemas externos europeos e internacionales, siempre dentro de los marcos legales aplicables.

En el ámbito europeo, la plataforma se integra con los sistemas y marcos siguientes:

- **Marco Prüm II**: para el intercambio automatizado de perfiles de ADN, datos dactiloscópicos y, en su caso, otros identificadores entre Estados Miembros de la UE.
- **EURODAC**: como base de datos dactiloscópica europea para solicitantes de asilo y determinadas categorías de entradas irregulares.
- **Sistema de Información Schengen (SIS RECAST)**: para la consulta y, cuando proceda, generación de señalamientos sobre personas y objetos.
- **Futuros sistemas del marco SMART BORDERS (EES, ETIAS)**, para facilitar la integración progresiva con los nuevos sistemas centralizados de gestión de fronteras exteriores de la UE.

A escala internacional y de cooperación policial ampliada, THOT prevé la interoperabilidad con:

- Las **bases de datos gestionadas por INTERPOL**, permitiendo verificaciones globales sobre identidades, documentos de viaje robados o perdidos y otros elementos de interés policial.
- El **Sistema de Información de EUROPOL (EIS)** para el intercambio y consulta de inteligencia criminal.

En el ámbito nacional, también se contemplan **intercambios con otras entidades clave**, como:

- **Organizaciones humanitarias (p.ej. Cruz Roja)**, en el contexto de Identificación de Víctimas de Desastres (IVD), para el intercambio controlado de datos Ante-Mortem y Post-Mortem.
- El **Instituto Nacional de Toxicología y Ciencias Forenses (INTCF)** y otros laboratorios o institutos de medicina legal, para la remisión y recepción de solicitudes y resultados.
- Los **Órganos Judiciales**, mediante la remisión telemática segura y con valor probatorio de informes y evidencias digitales.

Desde el punto de vista del Data Management Plan, el objetivo es que la entrada y salida internacional se gestione como un flujo más dentro de THOT/BINCIPO, y no como un circuito paralelo basado en correos electrónicos o procesos manuales.

### 9.5 Interoperabilidad entre Lotes

La **interoperabilidad entre el Lote 1 y el Lote 2** se articula mediante un Marco de Entendimiento específico, definido en la Fase I del proyecto y vigente durante todo su ciclo de vida.

Este marco tiene como objetivo asegurar que los sistemas de ambos lotes:

- Compartan un lenguaje de datos común (esquemas, identificadores, codificaciones).
- Empleen protocolos y patrones de integración coherentes.
- Se evolucionen de forma coordinada, sin romper la interoperabilidad alcanzada.

Para ello se establecen los siguientes instrumentos:

- **Comité de Interoperabilidad:** órgano en el que participan adjudicatarios de ambos lotes, responsable de acordar decisiones técnicas y funcionales que afecten a la interoperabilidad y de supervisar su cumplimiento.
- **Documento de Arquitectura de Interoperabilidad (DAI):** documento conjunto que define los modelos de datos compartidos, las interfaces expuestas por el Lote 1 y consumidas por el Lote 2, y las reglas de mapeo entre entidades (persona, asunto, vestigio, actividad, etc.).
- **Desarrollo conforme al DAI:** cualquier desviación respecto a la arquitectura acordada debe ser aprobada por el Comité de Interoperabilidad, evitando soluciones ad hoc no documentadas.
- **Plan de Pruebas de Interoperabilidad (PPI):** entregable específico que define el alcance, escenarios y criterios de aceptación de las pruebas.
- **Acuerdos de Nivel de Servicio (SLA) de interoperabilidad:** fijan compromisos concretos de disponibilidad, tiempos de respuesta y calidad de los servicios expuestos entre lotes.
- **Revisiones periódicas de interoperabilidad:** sesiones formales para verificar el cumplimiento del DAI y del PPI, identificar incidencias y acordar mejoras de cara a nuevas versiones.

En términos de gestión de datos, este marco garantiza que las decisiones de diseño que afectan a la semántica de los datos (nuevos campos, cambios de codificación, nuevos tipos de eventos) se aborden de forma coordinada, evitando divergencias entre el dato generado en escena (Lote 2) y el que se explota, analiza y custodia en el Lote 1.

## 10 METODOLOGÍA, CALIDAD Y GOBIERNO DEL DATO

### 10.1 Calidad del dato

Desde el punto de vista de **ISO 21043**, la calidad se consolida garantizando que cada manipulación relevante del vestigio queda registrada como **hecho trazable** con temporalidad, responsable y vínculo explícito con el vestigio afectado. Desde el punto de vista de **ISO/IEC 17025**, la calidad se consolida garantizando que el análisis se ejecuta bajo condiciones controladas, con personal cualificado, recursos trazables y resultados reproducibles y auditables; por ello, el modelo separa estrictamente **resultado** (medición/observación) de **afirmación de vinculación** (conclusión interpretativa), y exige invariantes semánticos sobre acreditación, requisitos de protocolo y consistencia temporal.

En términos prácticos, la calidad se implementa mediante un conjunto de controles complementarios que actúan en capas y se expresan como **contratos del modelo**:

1. **Calidad estructural (modelo y esquema).** Se aplican claves primarias estables (UUID), claves naturales cuando proceda (por ejemplo, código + version en PROTOCOLO), integridad referencial (FK), dominios controlados (CHECK) y restricciones temporales (p. ej., 'FechaFin' ≥ 'FechaInicio'). Estas medidas evitan estados inválidos “por construcción” y facilitan auditoría y explotación coherente.
2. **Calidad semántica (reglas operativas).** Se validan invariantes que no quedan capturados únicamente por cardinalidades. Entre las más relevantes: (i) la acreditación obtenida por un análisis se encuentra habilitada por al menos uno de los protocolos seguidos; (ii) las especialidades requeridas por protocolos quedan cubiertas por el conjunto de analistas participantes; (iii) los recursos requeridos por protocolo (reactivos/equipamiento) se encuentran entre los recursos usados; (iv) las afirmaciones de vinculación se sustentan en resultados existentes y respetan dominios controlados de confianza y calidad.
3. **Calidad probatoria (integridad y no repudio).** Para EVENTO, RESULTADO, AFIRMACION\_VINCULACION y referencias de objeto, se registra hash\_payload calculado sobre una canonicalización estable, y se correlaciona con el plano inmutable judicial. La verificación se ejecuta como procedimiento repetible: reconstrucción del payload canonicalizado, recálculo de huella y validación del anclaje. Para binarios críticos, el hash del fichero se conserva tanto en el canónico como en el plano inmutable, verificando que el objeto servido coincide con la evidencia registrada.
4. **Calidad de proceso (control de cambios y no conformidades).** Cuando se requiere corrección operativa, no se sustituye silenciosamente la evidencia registrada: el hecho original permanece y la rectificación se registra como nuevo hecho vinculado, preservando secuencia y huellas. En paralelo, se mantendrán mecanismos de revisión/aprobación y registro de incidencias (no conformidades) asociados a entidades operativas (EVENTO/ANALISIS/RESULTADO), de forma que el historial de correcciones y decisiones quede trazado sin erosionar la integridad probatoria.

Con este enfoque, la calidad del dato no depende de prácticas manuales aisladas, sino de un **diseño verificable**: restricciones en persistencia, validaciones semánticas en la capa de dominio y evidencia criptográfica independiente para custodia y auditoría. Esta base permite sostener la trazabilidad exigida por ISO 21043 y la robustez de registros y resultados alineada con ISO/IEC 17025, manteniendo coherencia con la arquitectura de almacenamiento y con el modelo de eventos descrito en el entregable.

## 10.2 Responsabilidades en el manejo de datos

A efectos de manejo y gestión de datos el Project Owner (CDTI) ejerce la autoridad de aprobación sobre la calidad y trazabilidad de la información en la validación de entregables por fase, en la evaluación de avances e hitos y en el informe final de fase I y transición a fase II, además de aprobar cambios de alcance o calendario cuando afecten a planes de datos. Por otro lado, complementa esa función como parte consultada en gestión de riesgos y cambios, en la revisión del DAI, en la aprobación del PPI y en el seguimiento DNSH, y se mantiene informado en validación de interoperabilidad y workshops.

El Project Manager (UTE) dirige la operación de datos del día a día: ejecuta la validación de entregables, el seguimiento DNSH y ético-legal, la evaluación de avances e hitos, el informe final y la comunicación del proyecto. Además, es el responsable último en la gestión de riesgos y cambios que inciden en modelos, flujos o políticas de datos; y actúa como consultado en interoperabilidad, DAI, PPI, cambios de alcance y workshops, asegurando coherencia entre equipos y comités.

El Solution Provider Team es el motor técnico del dato: asume la responsabilidad de producir, transformar y verificar información en la validación de entregables, la gestión de riesgos y cambios, la validación de interoperabilidad entre lotes, la revisión del DAI, la preparación del PPI y el seguimiento DNSH. Además, ostenta simultáneamente la responsabilidad operativa y última en la supervisión técnica de desarrollo e integración, y aporta criterio como consultado en la evaluación de avances, en el informe final y en los workshops, quedando informado cuando se aprueban cambios de alcance o calendario.

El Comité de Dirección aporta gobierno del dato desde la perspectiva estratégica: se le consulta para contrastar alineamiento y valor en validación de entregables, riesgos y cambios, evaluación de avances e hitos e informe final y asume la decisión final cuando los cambios de alcance o calendario impactan a los compromisos y métricas de información del proyecto.

El Comité de Interoperabilidad concentra la custodia de las reglas de intercambio y calidad semántica: es responsable último en la validación de interoperabilidad, en la revisión del DAI y en la aprobación del PPI, define criterios y evidencia de conformidad y actúa como consultado en validación de entregables, coordinación general, evaluación de avances e informe final, manteniéndose informado en riesgos y cambios, comunicación y workshops.

El Advisory Board vela por el buen gobierno del dato y su licitud: es responsable último del seguimiento del cumplimiento DNSH y de los aspectos éticos y legales y preside los workshops previos a hitos para contrastar enfoques, permaneciendo informado del resto de actividades para asesorar con visión externa cuando sea necesario.

La Policía Científica asegura la aplicabilidad operativa y la integridad probatoria de la información: participa como consultada en la definición de requisitos, en la validación de entregables, en la comunicación con impacto en datos y en la evaluación de avances, hitos e informe final, y se mantiene informada en interoperabilidad, DAI, PPI, riesgos y cambios, DNSH y workshops para preparar despliegues y aportar retroalimentación especializada en las siguientes fases.

La gestión de los datos en THOT es una responsabilidad compartida, con funciones claramente definidas para garantizar coherencia, trazabilidad y pleno cumplimiento normativo. Para ello, se designará:

- **Un Responsable de Gestión de Datos (Data Manager)**, que coordina este plan, mantiene el inventario y catálogo de conjuntos de datos, establece los estándares de documentación y versionado, y comprueba que los datos sean comprensibles y reutilizables por todos los socios.

- **Un Responsable de Ciberseguridad** (CISO del proyecto), que aplica el Plan Director de Seguridad de la solución THOT, traduciendo al proyecto los requisitos del Esquema Nacional de Seguridad y las mejores prácticas de ISO 27001/27701, incluyendo cifrado, copias de seguridad, gestión de claves, continuidad de negocio y respuesta a incidentes.
- **Un Delegado de Protección de Datos** (DPO), designado por la UTE, que en coordinación con los Delegados de Protección de Datos de las entidades participantes (incluida Policía Nacional cuando proceda) asegura la conformidad con el RGPD y la LOPDGDD, revisa la necesidad de Evaluaciones de Impacto en Protección de Datos (DPIA), valida cláusulas contractuales y controla el acceso de terceros a la información.
- **Un Responsable de Cumplimiento Ético**, que supervisa los aspectos éticos del manejo de datos y de los modelos de IA, evalúa riesgos de reidentificación y sesgos, revisa la explicabilidad y aprueba los planes de generación de datos sintéticos y anonimización/pseudonimización.

Los **responsables técnicos de cada módulo de la solución THOT** aplican las políticas de datos en su ámbito, incluyendo la minimización en origen, el etiquetado correcto, la generación de metadatos, el cifrado y la sincronización segura entre sistemas. Policía Nacional, como titular de los datos operativos, define y aprueba las condiciones de acceso, clasificación, conservación y difusión de cualquier dato real empleado en validaciones, así como los mecanismos de auditoría.

Finalmente, todos los datos se tratarán conforme a los estándares de calidad exigidos y respetarán la normativa vigente en materia de seguridad y protección de datos, de un modo que se desarrolla minuciosamente en el apartado 13.

### 10.3 Repositorios de confianza y principios FAIR

Las Directrices sobre la gestión de datos FAIR (Findable, Accessible, Interoperable & Reusable) establecen cuatro principios que rigen la gestión de los datos con el fin de que sean más localizables, accesibles, interoperables y reutilizables. En este sentido, THOT garantizará que todos los conjuntos de datos que se creen y procesen se gestionen de acuerdo con estas directrices, respetando en todo momento las restricciones legales y operativas propias del ámbito policial y forense y siguiendo el criterio **“tan abiertos como sea posible, tan restringidos como sea necesario”**.

Los datos que deban ser conservados y potencialmente compartidos, en especial aquellos con valor científico, demostrativo o reutilizable, se almacenarán en repositorios de confianza y se gestionarán mediante herramientas altamente contrastadas y reconocidas por la comunidad, que ofrezcan garantías de **seguridad, durabilidad, gobernanza clara y control de acceso**. Cuando la sensibilidad de los datos lo permita, se priorizará el uso de repositorios alineados con FAIR (p. ej., Zenodo o repositorios institucionales), garantizando la asignación de **identificadores persistentes** y páginas de aterrizaje con metadatos completos, siempre con la **autorización previa de PN**.

## 11 MAPEO DE DATOS POR DISCIPLINA FORENSE

Se añadirá en el próximo entregable

CONFIDENCIAL

## 12 MAPEO DE DATOS POR SERVICIO

Se añadirá en el próximo entregable

CONFIDENCIAL

## 13 POLÍTICA DE DATOS Y SEGURIDAD DEL DATO

### 13.1 Control de acceso y distribución

La información centralizada del proyecto se gestiona actualmente en el entorno **Microsoft OneDrive/SharePoint** perteneciente a Hi Iberia designada inicialmente como entidad responsable del **repositorio común de documentación del proyecto**.

Cada empresa participante conserva y gestiona la información que genera en sus propios sistemas corporativos, los cuales deberán cumplir con las políticas de seguridad y control definidas en el este Data Management Plan y en el Plan Director de Seguridad del proyecto, debiendo estar alineadas con las medidas del ENS (categoría alta, según el nivel de sensibilidad de la información), garantizándose así un nivel de protección homogéneo en toda la UTE.

Asimismo la UPV, como entidad subcontratada por la UTE, también establece un sistema de gestión y seguridad para los datos del proyecto que están alineados con las mismas medidas y categoría del ENS que el resto de las empresas participantes.

#### 13.1.1 Control de acceso

El acceso a la información se basa en el principio de privilegios mínimos y de “necesidad de conocer”. La autenticación de los usuarios se realiza mediante credenciales personales con autenticación multifactor (MFA), conforme a las políticas de seguridad propias del entorno de Microsoft 365 y las directrices del Esquema Nacional de Seguridad (ENS). Todos los accesos y modificaciones de información quedan registrados y auditados, garantizando la trazabilidad conforme a los requisitos del ENS.

#### 13.1.2 Transmisión y distribución de la información

La información entre las entidades de la UTE se intercambia a través de canales **cifrados (HTTPS/TLS 1.2 o superior)** dentro del entorno Microsoft 365 o mediante mecanismos equivalentes de **cifrado de extremo a extremo**, de acuerdo con los controles del ENS. Cualquier cesión de información a terceros requerirá autorización expresa del **Comité de Dirección de la UTE**, garantizando la confidencialidad mediante acuerdos de tratamiento o cláusulas contractuales equivalentes.

#### 13.1.3 Copias de seguridad y continuidad

El entorno SharePoint de HI IBERIA dispone de respaldo y redundancia geográfica nativa en la nube de Microsoft, conforme a las medidas de seguridad aplicables en su certificación ISO/IEC 27001.

Además, se establece una política de copias de seguridad complementarias, con periodicidad semanal o superior, en repositorios cifrados gestionados por los responsables de seguridad designados.

Los procedimientos de restauración serán verificados regularmente, en línea con los controles requisitos del ENS en materia de continuidad del servicio y recuperación ante desastres.

En caso de que se decida migrar el repositorio a una infraestructura independiente o a un entorno común gestionado por la UTE, se documentará el proceso garantizando la integridad y disponibilidad de los datos transferidos.

Se implementarán mecanismos de copia de seguridad inmutable basados en sistemas de almacenamiento WORM, que garanticen la inmutabilidad de la información, la resiliencia ante un ataque cibernético o por destrucción o alteración accidental, disminuyendo así los tiempos de RTO y ayudando, además, al cumplimiento normativo. Se establecerá un periodo para la persistencia de inmutabilidad en función de los distintos grados de criticidad de la información del proyecto.

### 13.2 Anonimización de datos personales

En los supuestos en que el proyecto THOT requiera el tratamiento de datos personales, las entidades integrantes de la UTE actuarán como corresponsables del tratamiento, de conformidad con el artículo 26 del RGPD.

Cada empresa garantizará el cumplimiento de los principios de licitud, lealtad, minimización y exactitud de los datos, así como la aplicación de medidas técnicas y organizativas adecuadas conforme al artículo 32 del RGPD, a la LOPDGDD. Los datos personales serán tratados únicamente para los fines previstos en el proyecto, antes de cualquier cesión, análisis o publicación de datos, se aplicarán técnicas de seudonimización o anonimización adecuadas para impedir la identificación de las personas físicas. Cuando se prevean tratamientos de riesgo elevado, se realizará una Evaluación de Impacto en la Protección de Datos (EIPD), coordinada por el Delegado de Protección de Datos (DPD) designado por la UTE o por la empresa miembro a la que le corresponda.

La UTE dispondrá de un procedimiento común de gestión de incidentes y brechas de seguridad, que incluirá la detección, registro, evaluación y notificación de incidentes tanto a la autoridad de control (AEPD) como a los interesados.

### 13.3 Análisis de vulnerabilidades

Se realizará un análisis sistemático de vulnerabilidades tanto a nivel de código como de infraestructura y de gestión de datos. En el plano del desarrollo software, se aplicarán **herramientas de análisis estático y dinámico de código** (por ejemplo, SonarQube, OWASP Dependency-Check, Snyk u otras soluciones equivalentes) para detectar librerías vulnerables, malas prácticas de programación, problemas de validación de entradas y errores de configuración de seguridad. En el plano de los datos, se utilizarán **herramientas de escaneo de configuraciones de bases de datos y almacenamiento** (por ejemplo, inspectores de permisos, revisores de cifrado en reposo y en tránsito, y analizadores de exposición de *endpoints*) que permitan identificar riesgos de acceso no autorizado o de filtración de información.

Los resultados de los análisis de vulnerabilidades se integrarán en el proceso general de gestión de riesgos de seguridad de la información, aplicando metodologías reconocidas (ej. MAGERIT) para determinar la probabilidad e impacto de cada vulnerabilidad. Se establecerá un registro de riesgos actualizado que permita priorizar las acciones correctivas en función del nivel de criticidad y del impacto sobre la confidencialidad, integridad, disponibilidad y trazabilidad de los datos.

Se incorporarán fuentes de ciber-inteligencia y alertas de vulnerabilidades nuevas y emergentes (por ejemplo, CVE, NIST NVD, INCIBE-CERT, CCN-CERT) para mantener actualizada la base de conocimiento de amenazas y anticipar medidas preventivas frente a vulnerabilidades críticas recién descubiertas.

Para asegurar la capacidad de restaurar la disponibilidad y el acceso a los datos de forma rápida en caso de incidente físico o técnico, el análisis de vulnerabilidades se complementará con la revisión periódica de los procedimientos de copia de seguridad, recuperación ante desastres y alta disponibilidad. Se verificarán los **tiempos de recuperación (RTO)** y los **puntos de recuperación (RPO)** definidos, y se evaluará su adecuación al nivel de criticidad de los datos tratados, ejecutando pruebas controladas de restauración y conmutación por error.

Finalmente, se establecerán procesos de verificación, evaluación y revisión regulares. Estos incluirán **revisiones de seguridad en cada versión relevante del software**, auditorías técnicas programadas, **reevaluaciones de riesgos** cuando se introduzcan cambios significativos en la arquitectura o en los tratamientos de datos, y la **actualización continua del inventario de vulnerabilidades conocidas**. Los resultados de estos análisis se documentarán y se utilizarán para priorizar medidas correctivas, registrar evidencias de cumplimiento y

alimentar un ciclo de mejora continua en la gestión de la seguridad de la información dentro del Data Management Plan.

Se adoptará un enfoque DevSecOps, integrando los controles de seguridad y las herramientas de análisis en el propio ciclo de desarrollo y despliegue continuo (CI/CD). Las pruebas de seguridad se automatizarán en las fases de integración y entrega, garantizando que ningún componente vulnerable pase a producción sin revisión previa.

#### 13.4 Política de retención de los datos

Los datos se conservarán durante la **vigencia del proyecto** y el periodo necesario para cumplir con las **obligaciones legales, contractuales y de auditoría** derivadas del mismo (habitualmente entre 5 y 10 años, según la normativa correspondiente).

La política de conservación y eliminación se ajustará a los principios de **limitación de conservación** del RGPD y a los controles del ENS.

La política de retención será revisada anualmente por el **responsable de seguridad del proyecto** y validada por el **Comité de Dirección de la UTE**, conforme al principio de **mejora continua (PDCA)** establecido por la **ISO/IEC 27001**.

Los conjuntos de datos sintéticos o anonimizados que se publiquen se preservarán durante la vida del proyecto y, al menos, cinco años más para favorecer su reutilización.

Los conjuntos restringidos utilizados en I+D y pruebas internas se conservarán durante la vida del proyecto y, al menos, tres años adicionales, salvo que Policía Nacional indique un periodo distinto.

Los registros de auditoría y las evidencias de cadena de custodia generadas para pruebas se conservarán un mínimo de cinco años. Si se han empleado datos reales en validaciones, su retención y destrucción se harán conforme a las reglas que marque Policía Nacional.

Los datos se almacenarán en infraestructuras certificadas (Microsoft 365) ubicadas dentro del Espacio Económico Europeo y con garantías de cumplimiento del ENS e ISO/IEC 27001.

Los datos y copias de respaldo se conservarán cifrados (AES-256), con control de acceso basado en roles y registro de todas las operaciones.

Cuando un conjunto de datos haya llegado al final de su vida útil, se destruirá aplicando métodos verificables que imposibilitan su recuperación, generando un acta de destrucción que queda en el registro.

## 14 MIGRACIÓN Y EVOLUCIÓN DEL DATO

### 14.1 Migración inicial

Se añadirá en el próximo entregable

CONFIDENCIAL

## 15 ASPECTOS ÉTICOS

### 15.1 Recogida y empleo de los datos

El proyecto da prioridad absoluta a trabajar con datos anonimizados o sintéticos y solo empleará datos reales si no hay otra opción para lograr una validación fiable y siempre con autorización expresa de Policía Nacional.

Cuando participen personas voluntarias en pruebas de usabilidad u otras actividades similares, se informará con claridad del propósito, de los datos que se recogerán y de sus derechos, y se recabará su consentimiento cuando sea pertinente.

La recogida de datos en THOT se realiza exclusivamente con fines legítimos de Policía Científica y bajo autoridad competente, respetando los principios de necesidad y proporcionalidad.

### 15.2 Sensibilidad, privacidad y protección de los derechos humanos

THOT clasifica los conjuntos en tres niveles (Público, Restringido y ConfidencialPN) y aplica controles acordes con su sensibilidad, prestando especial atención a categorías que pueden afectar de forma intensa a la privacidad y a los derechos fundamentales, como biométricos, genéticos, dentales, geolocalizaciones finas, imágenes de víctimas o de menores, y contenidos que muestren técnicas operativas.

Los datos de I+D y de formación serán sintéticos o anonimizados; cuando sea imprescindible conservar relaciones para reproducibilidad, se recurrirá a seudonimización con custodia separada de claves y control de accesos estricto, sabiendo que la seudonimización no saca a los datos del ámbito de protección.

Se realizarán Evaluaciones de Impacto en Protección de Datos (DPIA) o su análisis equivalente en el marco de LO 7/2021 (transposición de la Directiva 2016/680) para tratamientos que lo requieran, como el uso de analítica de IA en apoyo a decisiones operativas. Estas evaluaciones considerarán, además del RGPD/LOPDGDD y la LO 7/2021, el Esquema Nacional de Seguridad, el estado del arte y el riesgo de reidentificación, documentando medidas de mitigación, plazos de retención y criterios de acceso.

El respeto a los derechos humanos (privacidad, no discriminación, presunción de inocencia y libertad de expresión) guía la configuración de los servicios de IA., tales como las funciones de recomendación, alerta y generación de informes operan como apoyo a la toma de decisiones, con supervisión humana obligatoria, explicaciones comprensibles y registro de razones.

No se adoptarán decisiones automatizadas con efectos jurídicos o impacto similar sin revisión humana.

Cualquier capacidad de inferencia sensible (por ejemplo, detección de emoción en audio) se deshabilita por defecto y solo puede activarse con autorización expresa, con finalidad acotada, señalización al usuario y sin uso para atribuir culpabilidad o realizar perfiles indebidos.

Los modelos de inteligencia artificial se revisan para detectar posibles sesgos que pudieran producir resultados discriminatorios o injustos, y se adoptan medidas mitigadoras para reducir estos riesgos.

### 15.3 Responsabilidad, transparencia y supervisión

THOT adopta un modelo de gobernanza basado en revisiones internas continuas y “ética por diseño”. Este enfoque se articulará como un ciclo de revisión integrado en el desarrollo y el despliegue de los servicios, con responsabilidades claras y puntos de control en los hitos del proyecto.

Se designará un **Responsable de Cumplimiento Ético** del Proyecto que coordinará las revisiones y mantendrá el registro de riesgos éticos, con el apoyo del DPO, que actuarán en calidad consultiva cuando un tratamiento pueda implicar riesgos elevados.

Este modelo no elimina la supervisión de Policía Nacional: cualquier cuestión sustantiva se eleva para su validación y, en caso de riesgos significativos, se prevé la convocatoria ad hoc de una revisión con expertos de PN antes de avanzar.

La transparencia se asegura mediante documentación proporcional y comprensible del porqué y para qué de cada función sensible, especialmente en los servicios de THOT que traten con datos sensibles y de gran importancia para la policía científica (generación automatizada de informes, recomendaciones y alertas, multilinguaje, cadena de custodi).

Se elaborarán “data cards” (descripciones claras de los conjuntos de datos) para los conjuntos utilizados, “model cards” (descripciones claras de los modelos de IA) para los algoritmos desplegados en el borde con su dominio de validez, métricas y salvaguardas, y bitácoras de decisiones que recogen alternativas evaluadas, criterios éticos aplicados y medidas de mitigación.

La plataforma registrará de forma auditable capturas, transformaciones, accesos, sincronizaciones y decisiones relevantes, lo que permite reconstruir la actuación y verificar que las recomendaciones de IA han sido revisadas por un agente antes de generar efectos operativos.

Para prevenir el uso indebido, el diseño y la configuración por defecto limitan la funcionalidad a lo necesario y mantienen deshabilitadas capacidades de mayor riesgo hasta su habilitación explícita tras revisión (por ejemplo, comparaciones 1:N más allá de listas locales autorizadas).

Se aplicarán controles de acceso por roles y principio de mínimo privilegio, autenticación multifactor, bloqueo de exportaciones fuera de canales aprobados y políticas claras de casos de uso prohibidos en la UI.

El proyecto promoverá internamente una formación inicial y recordatorios periódicos en ética, protección de datos y uso responsable de IA. Este enfoque se alinea con el marco ALTAI mediante autoevaluaciones y con la legislación aplicable (RGPD/LO 7/2021 y ENS), activando cuando corresponda una evaluación de impacto en protección de datos antes de introducir nuevas funciones o modificar sustancialmente las existentes.

## 16 CONCLUSIONES

CONFIDENCIAL

## ANEXOS

CONFIDENCIAL