UNIVERSIDAD AUTÓNOMA DE MADRID

ESCUELA POLITÉCNICA SUPERIOR

TRABAJO FIN DE MÁSTER

Metodología forense para el desarrollo de un proyecto eDiscovery en un entorno profesional

AUTOR: Elena Ortega Herreros

TUTOR: Álvaro Ortigosa Juárez

Enero 2017

Metodología forense para el desarrollo de un proyecto eDiscovery en un entorno profesional

AUTOR: Elena Ortega Herreros
TUTOR: Álvaro Ortigosa Juárez

Dpto. de Ingeniera Informática
Escuela Politécnica Superior
Universidad Autónoma de Madrid
Enero 2017

Resumen

La globalización de la información hace que cada vez se almacene más información digitalmente frente al tradicional papel. También el uso creciente del correo electrónico y la mensajería instantánea como Whatsapp, Hangouts o Skype sustituye a las comunicaciones verbales. Uniendo estas nuevas formas de creación de información y comunicación al mundo laboral surgen multitud de investigaciones que centran su foco en el análisis de la información digital.

Del proceso de analizar esta información de forma correcta y exhaustiva nace la necesidad de disponer de guías, como la que este proyecto desarrolla, para cumplir internacionalmente los estándares mínimos al llevar a cabo una investigación. Disponer de metodologías y guías de buenas prácticas en estas investigaciones de naturaleza tan reciente es indispensable para formar a nuevos profesionales en este campo y poder cubrir la creciente demanda, así como ayudar a salvaguardar sus responsabilidades derivadas de estas investigaciones.

Este proyecto cubre todas las fases del *Electronic Discovery Reference Model* a la vez que sigue las recomendaciones de la ISO/IEC 27037:2012, para desarrollar una guía que permita llevar un correcto seguimiento de las investigaciones informático-forenses. Además, incluye un ejemplo práctico utilizando las tecnologías más utilizadas actualmente en el mundo profesional.

Palabras clave: Análisis Forense, eDiscovery, Investigación, Evidencia, Análisis de Datos.

Abstract

The globalization of information means there is more data stored in digital form versus the traditional paper format. The increasing usage of email and instant messaging such as Whatsapp, Hangout or Skype are also replacing verbal communication. Combining this new forms of creating and communicating information in the existing corporate world, creates a necessity of having many forensic investigations with digital information analysis as focus.

From the process of analyzing this information in a correct and exhaustive manner has borne the necessity of having guides, such as the following developed in this project, at people's disposal so as to comply with a minimum set of international standards when conducting investigations. The availability of best practice guides in a field so recent is essential to meet the increased demand of newly trained professionals, as well as to help protect their export witness responsibilities that may arise during the investigations.

This project covers all the steps in the Electronic Discovery Reference Model whilst following the recommendations of the ISO/IEC 27037:2012 in order to create a guide to conduct a proper digital forensic investigation. It also includes a practical example using the most popular technologies currently available in the field.

Keywords: Forensics, eDiscovery, Investigation, Evidence, Data Analisis.

Agradecimientos

Quiero agradecer este proyecto a mi tutor, Álvaro, por darme espacio y confianza para realizar este proyecto. Por su apoyo y consejo constante.

A Daniel por no preguntar, ¿cómo va el TFM? y apoyarme incondicionalmente. Por confiar en mi capacidad y estar siempre a mi lado.

A mi compañero Jaime, por darme soporte con su experiencia y por su infinita paciencia escuchando mis dudas y conclusiones sobre nuestro trabajo.

Y no por última menos importante, a Natalia, por ser la mejor compañera de viaje. La mejor compañera durante estos 8 años de incesantes altibajos, pero que por fin, hemos conseguido superar satisfactoriamente.

Índice

Índice de figuras

[1] http://www.edrm.net

Índice de tablas

1. Introducción

La forma de trabajar ha cambiado, ya no se concibe no utilizar los programas de ofimática para redactar un documento o crear una presentación. El correo electrónico y la mensajería instantánea son herramientas utilizadas a diario en el mundo tanto personal como empresarial. En repetidas ocasiones ambos mundos conviven en un mismo dispositivo y los usuarios no realizan una diferenciación de uso entre sus dispositivos corporativos y personales.

Estos dispositivos guardan una gran cantidad de información empresarial y personal, y son la clave de las actuales investigaciones informático-forenses realizadas a raíz de demandas, defensas o simples sospechas sobre el uso incorrecto de los dispositivos empresariales por parte de los empleados. En muchas ocasiones se utilizan los dispositivos electrónicos proporcionados por la empresa para realizar fraude, competencia desleal, extraer información privilegiada, etc. Estas prácticas son cada vez más frecuentes y es necesario disponer de profesionales formados para realizar investigaciones y tratar adecuadamente esta información electrónica que dará lugar a pruebas en litigios y conciliaciones empresariales.

El contexto de estas investigaciones es riguroso y requiere de una metodología estricta que asegure el correcto tratamiento de la información y asegure la no manipulación de la misma. Los profesionales que desempeñan esta labor se denominan peritos informáticos, y salvaguardar lo máximo posible las responsabilidades derivadas de su actividad es un necesidad creciente.

De otro lado, la cantidad de información a procesar en una investigación se vuelve compleja y en ocasiones inviable de revisar de forma tradicional. Nos encontramos con el reto del tratamiento masivo de datos, tan actual e importante a la hora de ser ágiles y eficientes encontrando evidencias que puedan dilucidar los casos de investigación. Tener claros los objetivos y elegir las herramientas adecuadas para lograrlos es indispensable para agilizar los procesos y llegar a conclusiones en un tiempo razonable y con los recursos disponibles.

1.1. Motivación y objetivos

Crear buenas prácticas para salvaguardar la responsabilidad de los peritos informáticos ha sido la motivación principal para el desarrollo de este proyecto, así como la necesidad de agrupar procedimientos y disponer de una guía básica para actuar ante investigaciones informático forenses. El principal objetivo es aunar las distintas fuentes disponibles, como la ISO/IEC 27037:2012, el *Electronic Discovery Reference Model*, así como escoger las buenas prácticas de la experiencia en este ámbito y corregir las malas prácticas detectadas. Con esta información se pretende crear una guía de carácter práctico y con capacidad de ser utilizada, sobre todo en las primeras fases de Identificación, Preservación y Adquisición, desde el primer momento. Por las características del resto de las fases, se ha centrado la investigación en un tipo concreto de investigación. Este proyecto no pretende cubrir todas las tipologías de casos, pues sería imposible dadas las diferentes características entre proyectos, pero sí ser una base que permita ser modificada y adaptada a las necesidades de las investigaciones así como al grupo de investigadores.

Como se verá en las siguientes secciones la guía se basa en el diseño de una serie de formularios con el fin de facilitar la anotación de información necesaria y estipular los pasos a seguir a los peritos informáticos. Estos formularios son de elaboración propia y se han diseñado a lo largo de diversos proyectos reales que han ido marcando las diferentes necesidades.

Además, se pretende crear un pequeño caso que muestre cómo poner en práctica la guía y cómo adaptarla a las necesidades, así como mostrar las tecnologías más usadas en el mundo empresarial.

1.2. Estructuración del documento

El presente documento se estructura en dos grandes secciones, la guía teórica y la ilustración de la misma a través de un pequeño caso práctico. Respecto a la primera sección, se han trasladado a los anexos todos los formularios completos para facilitar la lectura, pero no por ello son menos importantes, al contrario, contienen la mayor parte de información sobre las primeras fases del *Electronic Discovery Reference Model*.

Sobre la segunda sección, se ha dividido a su vez en dos. Esta primera subsección, dada la magnitud y viabilidad del proyecto expuesto, se ha desarrollado en un marco ficticio con un dispositivo personal trabajando así, sólo con una pequeña parte de lo expuesto en la sección teórica. Para la segunda subsección, gracias a la tecnología usada, se ha escogido gran parte de la información disponible sobre el caso de estudio elegido, para mostrar la magnitud de la investigación y el tratamiento masivo de datos del que venimos hablando.

2. Contexto

Una vez se ha concluido con el cliente que la información que puede esclarecer el caso de la investigación en curso se cierne sobre dispositivos digitales y la información almacenada en los mismos, se debe comenzar dando cumplimiento a la ISO/IEC 27037:2012 y siguiendo todos los pasos del *Electronic Discovery Reference Model*.

El estándar ISO/IEC 27037:2012 [2] define las" *Directrices para la identificacion, recolección, adquisición y preservación de la evidencia.*" en un marco internacional. Este estándar define tres principios que debe cumplir toda evidencia digital: la relevancia, la confiabilidad y la suficiencia; así como los roles de los actores que deben velar por el cumplimiento de los mismos: el *Digital Evidence First Responder* (DEFR) y el *Digital Evidence Specialist* (DES).

La **relevancia** hace referencia a la importancia de la información obtenida de la evidencia respecto al caso investigado.

La **confiabilidad** es la premisa más importante para que la evidencia sea aceptada como prueba de una investigación. Este principio implica que los resultados deben ser reproducibles por un tercero siguiendo el mismo proceso.

El principio de **suficiencia** indica que la evidencia debe sustentar por sí misma la conclusión de la investigación.

Durante una investigación hablaremos siempre de evidencia potencial hasta demostrar que cumple los tres principios anteriormente citados.

Para que estos principios puedan ser llevados a cabo es necesario la acción de los roles DEFR y DES. El **DEFR** es la primera persona que llega al escenario y se encarga de documentar el estado de la evidencia y preservarlo hasta que llegue el segundo actor. El **DES**, puede realizar las mismas acciones que el DEFR pero además posee los conocimientos técnicos para llevar a cabo la recolección de la evidencia.

Es importante que los DEFR y DES sean profesionales independientes y no estén involucrados en la investigación. En España se les denomina *peritos informáticos*. En el transcurso de su trabajo es necesario salvaguardar las consecuencias que pudieran tener derivadas de sus acciones, por lo que se debe especificar y documentar todo el proceso para minimizar el riesgo de contaminar la evidencia y su posterior pérdida de condición en términos legales.

Para ello, este trabajo se va a basar en la guía de referencia *Electronic Discovery Reference Model* (EDRM), definiendo cada una de las etapas en procedimientos concretos y protocolos a seguir en la investigación sobre una evidencia digital.

3. Modelo Electronic Discovery Reference Model (EDRM)

El modelo *Electronic Discovery Reference Model*, en adelante EDRM, da respuesta a la necesidad de enumerar y definir las prácticas realizadas sobre evidencias digitales a nivel global. El EDRM divide el proceso en 5 etapas como se muestra en la imagen 1 y establece las guías que deben seguirse para equiparar dentro del mismo marco todas las investigaciones.

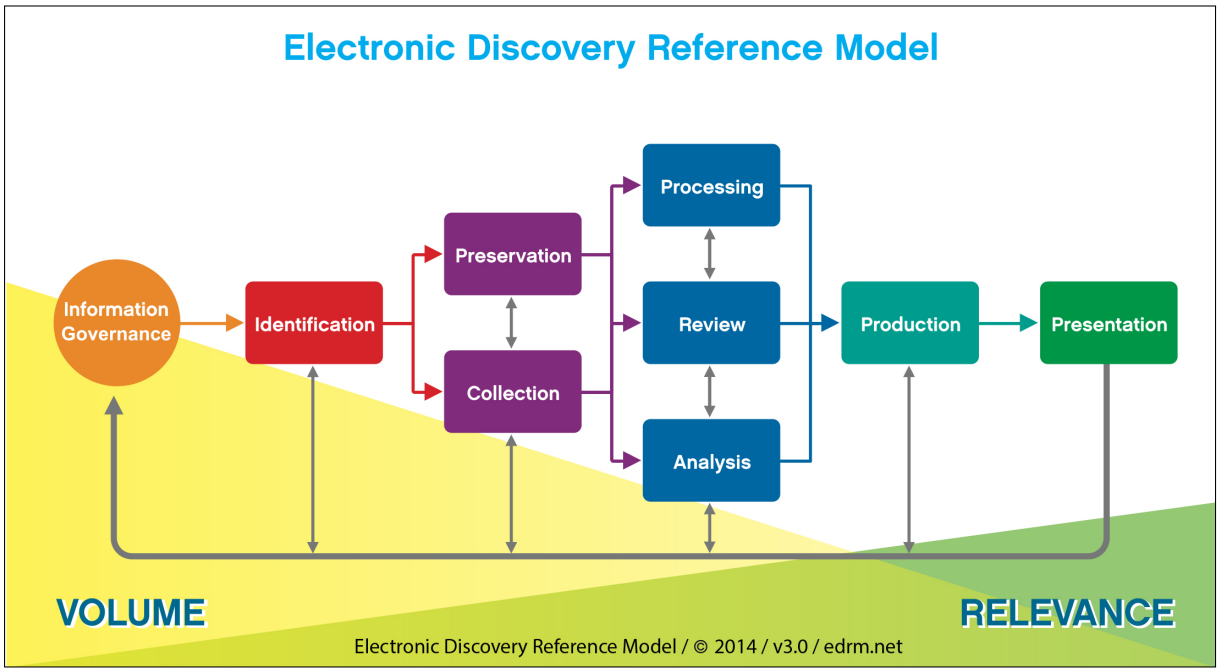


Figura 1: Diagrama ilustrativo del Electronical Discovery Reference Model[2]

En las siguientes subsecciones se va a definir cada una de las etapas y a modelar los procedimientos necesarios a seguir para su correcto desarrollo. La metodología de trabajo se ha desarrollado en torno a una serie de formularios que estructuran y recogen la toda información necesaria.

3.1. Identificación

La primera fase es la Identificación tanto del caso como de las evidencias que lo componen. En esta fase se debe acotar la investigación y determinar qué es relevante y por lo tanto, qué es necesario que pase a formar parte como potencial evidencia.

En primer lugar se debe establecer un equipo de trabajo formado por personas de la compañía contratante y el equipo forense, que permita la comunicación y la gestión del proyecto de investigación. La compañía investigada debe facilitar la información solicitada y la correcta colaboración de aquellas personas clave en la investigación.

En las siguientes subsecciones definiremos los pasos a dar para definir dónde y cómo se encuentra la información relevante.

[2]http://www.edrm.net

3.1.1. Gestión de información

Para comenzar, se debe recopilar información sobre la existencia, o no, de las políticas de gestión de la información de la empresa para poder definir los límites de los datos y tener una visión de cómo se estructura y custodia la información internamente. Para ello, se ha definido el formulario *Gestión de información* [Formulario A.1].

Este formulario recoge información sobre la *Política de Gestión de la Información*, el *Protocolo de Retención de Información* y una primera suposición de dónde se podría encontrar información relevante para la investigación.

En casos en los que el número de investigados, de ahora en adelante denominados *custodians*, sea considerable, será necesario definir un **Perímetro de Custodians** que permita establecer los límites de la investigación y facilitar la logística temporal de la misma. Es importante disponer de los siguientes datos de un *Custodian* para poder llevar un seguimiento: *Identificador único* propio del proyecto a cada empleado además del asignado por la compañía, así como información que nos permita establecer contacto con ellos como el *Correo Electrónico* o el *Número de Teléfono*, su *Cargo* y *Departamento* dentro de la compañía.

3.1.2. Departamento Tecnologías de la Información

Tras obtener una primera visión sobre la política de gestión de la información en la empresa, continuaremos con el lado más técnico a través del Departamento de Tecnologías de la Información o su homólogo. Para ello, se ha diseñado el formulario *Departamento IT* [Formulario A.2] que pregunta por la gestión de servidores, buzones de correo; la información almacenada en red, ordenadores corporativos y bases de datos.

3.1.3. Entrevistas

Tras los dos pasos anteriores, ya estamos preparados para empezar a realizar entrevistas a los *custodians* del perímetro e identificar y enumerar los dispositivos físicos de los empleados. Esta tarea de identificación puede llevarla a cabo un responsable de la empresa contratante, por ejemplo, a través del inventario de la empresa. En el caso de que sea el equipo forense el encargado de identificar y recolectar los dispositivos de un custodian será necesario una planificación entre el equipo de entrevistas y el equipo de adquisiciones que permita disponer de tiempo para llevar a cabo todo el proceso y hacérselo saber al empleado.

Para estandarizar lo máximo posible las entrevistas se ha elaborado el formulario *Entrevista Custodian* [Formulario A.3]. Este breve formulario pretende identificar todos los dispositivos corporativos del empleado, dejar constancia sobre su conocimiento sobre el proceso llevado a cabo y las retenciones legales. También recoge si el usuario posee carpetas personales dentro de estos dispositivos, aspecto muy importante a la hora de dilucidar el contexto en el que se produjo la evidencia en caso de hallarse en este área personal, y determinar si es válida en el proceso.

Sobre la información almacenada en red y en el servidor de correo electrónico, es el Departamento de Tecnologías de la Información el encargado de facilitar los accesos a los servidores y las rutas correspondientes de cada usuario.

3.2. Preservación y Adquisición

En el diseño de este segundo paso especificado en el EDRM, se ha combinado la preservación y la adquisición a través de la Cadena de Custodia, con la finalidad de que en cada proceso se rellene la información generada y que al finalizar la cadena ésta recoja toda la información necesaria sobre la preservación y adquisición de la potencial evidencia. La Cadena de Custodia, en adelante CoC, es un elemento fundamental para garantizar el principio de confiabilidad.

3.2.1. Preservación

La CoC debe recoger quién posee la potencial evidencia en cada momento desde que la entrega el usuario hasta que se devuelve al mismo. Cada dispositivo debe tener su propia CoC que recoja tanto los datos del dispositivo para ser identificado como quién la preserva en cada momento.

En la especificación de la custodia del dispositivo es imprescindible que aparezca quién entrega a quién el dispositivo, así como su firma y la fecha de la acción. El firmante debe hacerse cargo del dispositivo en el periodo de tiempo que esté en su posesión. Figura 2.

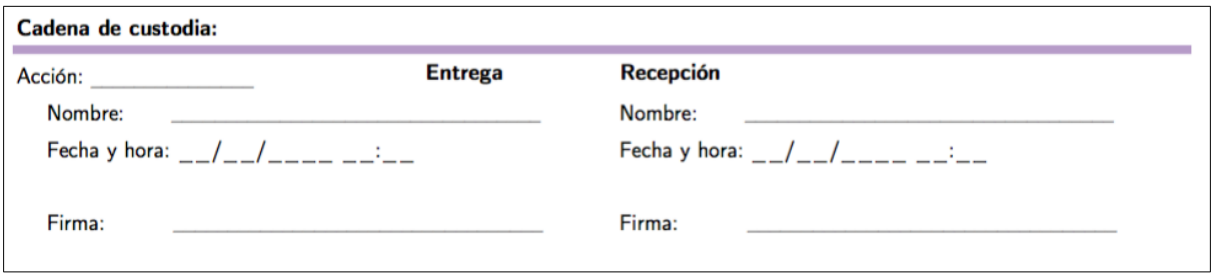


Figura 2: Custodia.

Los datos imprescindibles para identificar un dispositivo son:

- Identificador único de la evidencia.
- Identificador único del dispositivo. (Normalmente asociado con el número de inventario de la empresa.)
- Datos identificativos del custodian.
- Datos identificativos del dispositivo (Descripción, S/N...).

Respecto al identificador único de la evidencia e identificador único del dispositivo se debe tener en cuenta la siguiente premisa: Un dispositivo puede tener asociadas más de una evidencia si es adquirido en distintas ocasiones. Cada vez que un dispositivo pasa de la custodia del propietario al equipo forense se debe crear una nueva evidencia. El identificador del dispositivo no varía en ninguno de los casos.

En la fase de identificación se deben anotar todos los datos que permitan identificar posteriormente el dispositivo. Dadas las diferencias entre dispositivos, se han diseñado los siguientes formularios:

- **Dispositivo compuesto:** Un dispositivo compuesto es aquel que en su interior contiene otro. Por ejemplo, un ordenador, el cual dispone de datos propios y de un disco interno identificable. [Formulario A.4]

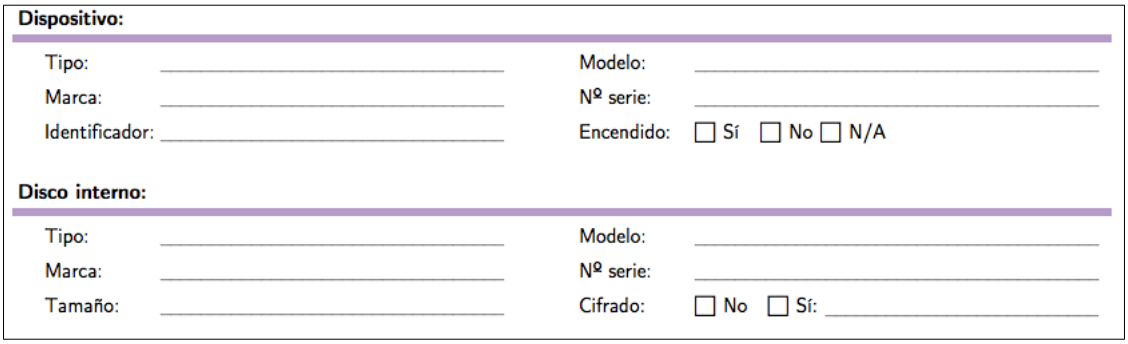


Figura 3: Información sobre el dispositivo compuesto.

- **Móvil / Tablet:** Los móviles y tablets son dispositivos particulares que contienen, normalmente, tarjetas SIM y ocasionalmente memorias internas. [Formulario A.5][3] Para las tarjetas de memoria se podría adjuntar el Formulario A.7 o dejar constancia a través de las notas.

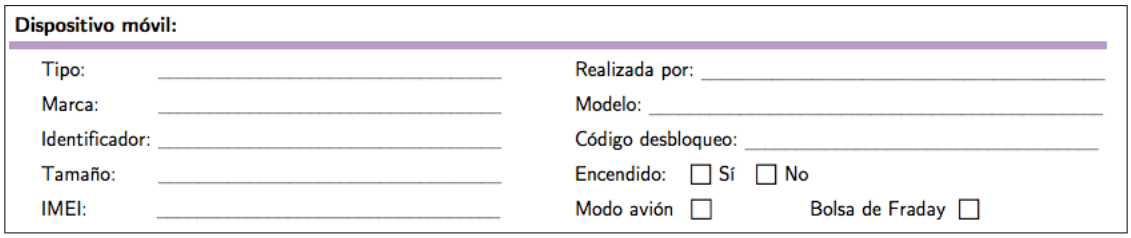


Figura 4: Información sobre el dispositivo móvil.

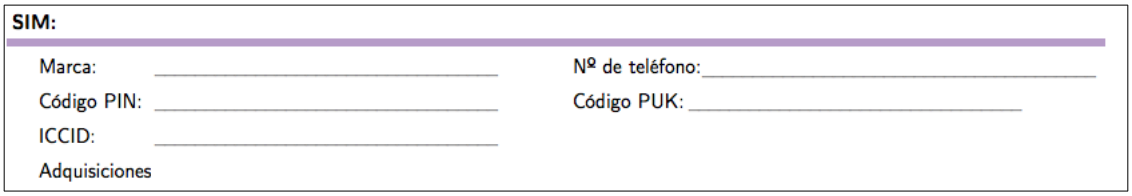


Figura 5: Información sobre la tarjeta SIM.

- **Extracción de información:** Este formulario es específico para adquisiciones de servidores o cloud, como por ejemplo una extracción de Google Vault o unas carpetas de red. [Formulario A.6]

[3]Este formulario se ha diseñado para realizar adquisiciones de dispositivos móviles con Cellebrite UFED, uno de los softwares y hardwares más extendidos y potentes a la hora de realizar adquisiciones de dispositivos móviles. [4]

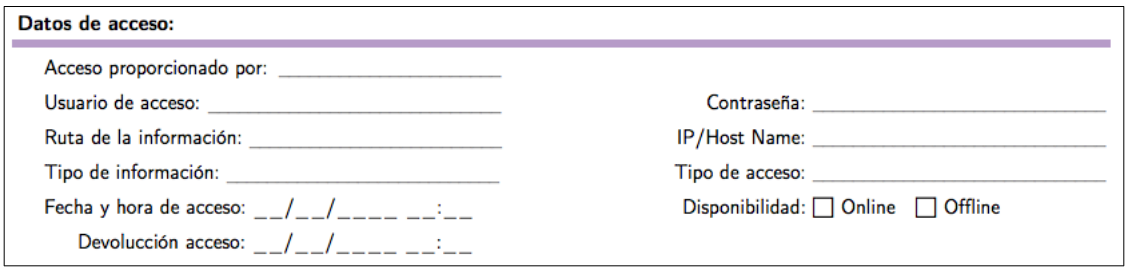


Figura 6: Información sobre extracción de información de servidores.

- **Otros dispositivos:** Para todos aquellos dispositivos más simples o no descritos en los formularios anteriores, como por ejemplo CDs o tarjetas de memoria. [Formulario A.7]

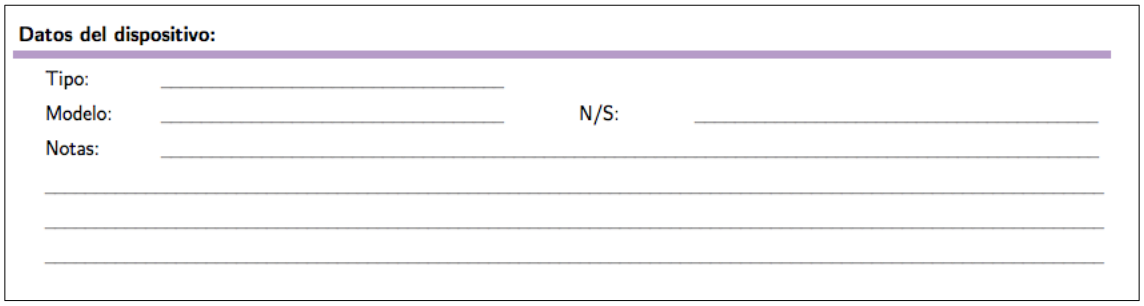


Figura 7: Información sobre un dispositivo simple.

3.2.2. Adquisición

Una vez tenemos el dispositivo identificado y en qué estado se encuentra deberemos proceder a prepararlo para adquirirlo. En todo momento se va a hablar sobre adquisiciones de la memoria física, dejando como trabajo futuro la memoria volátil por su complejidad y su finalidad en este tipo de casos. Antes de empezar es importante tener en cuenta las siguientes premisas.

- El dispositivo, no móvil, debe encontrase apagado. El DEFR debe asegurarse que el custodian se lo entrega en este estado. En el caso de encontrarse encendido deberá apagarse quitando la fuente de alimentación del mismo o forzando el apagado a través del botón correspondiente. Por seguridad, es muy recomendable extraer siempre la batería de los dispositivos portátiles antes de manipularlos. Con estas acciones evitamos alterar el sistema y los registros del mismo.

- Si nos encontramos con un dispositivo móvil, deberá estar en modo avión para evitar accesos remotos al mismo. En caso de encontrarse apagado y desconocer si el dispositivo se encuentra en este modo, deberá ser introducido en una bolsa de *Faraday* y encenderlo desde dentro para seleccionar el modo avión.

Los tipos de adquisiciones más habituales que nos podemos encontrar son las **Adquisiciones Físicas** y las **Adquisiciones Lógicas** de memorias. Siempre que el dispositivo lo permita debe realizarse una adquisición física puesto que proporciona la información total del dispositivo a copiar. En casos como las carpetas de red sólo se podrán realizar adquisiciones lógicas si el acceso es remoto o la información a copiar no es la totalidad del disco.

Para todos los tipos de adquisiciones se debe recoger la siguiente información con el fin de asegurar y preservar el principio de confiabilidad. Figura 8. Es importante destacar el papel de los hashes en el proceso, estos son vitales a la hora de garantizar este principio. Si un tercero reproduce el proceso deberá obtener los mismos hashes obtenidos en la primera adquisición. En el caso de verse alterada mínimamente la información, al readquirirla bajo las mismas premisas anteriores, los hashes cambiarían totalmente y pondrían de manifiesto que se ha producido una manipulación sobre los datos. Además, se deberá recoger la información sobre dónde se almacena la evidencia y su seguridad. Como se ha mencionado anteriormente, esta información se recogerá junto a la CoC y los datos del dispositivo, quedando documentada conjuntamente la preservación y adquisición de cada dispositivo.

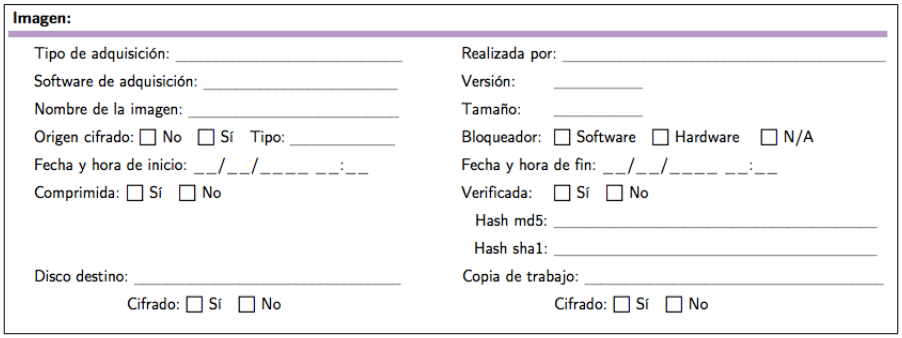


Figura 8: Información sobre la imagen.

En el caso particular de las adquisiciones de dispositivos móviles los tipos de adquisiciones son más amplios y su disponibilidad dependerá del modelo. En la figura 9 podemos ver los tipos de adquisiciones que se pueden realizar con Cellebrite UFED [4].

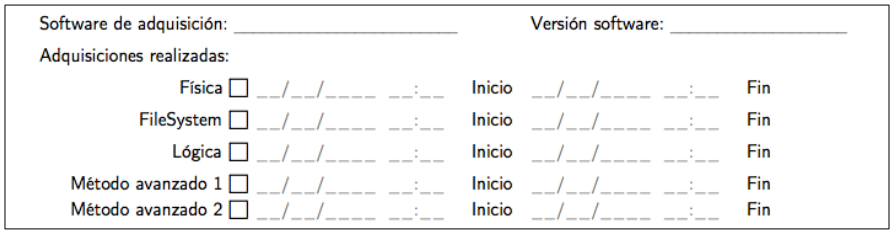


Figura 9: Información a anotar obtenida de Cellebrite UFED para adquisiciones de dispositivos móviles.

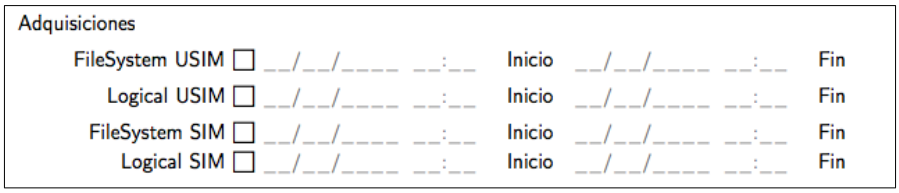


Figura 10: Información a anotar obtenida de Cellebrite UFED para adquisiciones de tarjetas SIM

Tras realizar adquisiciones con UFED se debe realizar una adquisición lógica de la información obtenida para encapsularla dado que Cellebrite UFED no ofrece este formato. Con esta acción garantizamos la no alteración de la información.

Otro aspecto, no menos importante, a tener en cuenta cuando se realiza una adquisición es la importancia de no alterar las evidencias. Al igual que es importante mantener un dispositivo móvil en modo avión para evitar posibles accesos remotos, es importante bloquear físicamente contra escritura los dispositivos que se conectan a la estación de trabajo con el fin de que ésta no pueda realizar cambios en ellos.

Una buena práctica para complementar estos formularios es realizar fotografías a los dispositivos identificando la evidencia y mostrando los datos especificados en el formulario.

3.3. Procesamiento, Revisión y Análisis

Llegados a este punto no necesariamente se necesita al DEFR, ni se espera que disponga de los conocimientos necesarios para llevar a cabo el procesamiento, revisión y análisis. Un técnico cualificado en estos ámbitos puede seguir con la investigación dado que se debe trabajar con la copia de trabajo y no con los dispositivos originales.

3.3.1. Procesamiento

Como en todas las etapas, no se debe dejar de lado el cumplimiento del principio de confiabilidad. Para ello, seguiremos sirviéndonos de formularios [Formulario A.8] para dejar constancia de todas las acciones realizadas y los resultados obtenidos.

El objetivo del procesamiento es identificar los archivos a revisar mediante el desglose, la normalización y el filtrado de archivos. Para ello es necesario seguir los siguientes pasos: Evaluar los datos, Prepararlos, Seleccionarlos, Normalizarlos, Validarlos y Exportarlos. Estos pasos no necesariamente se deben aplicar linealmente, estará en función de las plataformas utilizadas y se deberán adaptar a las necesidades.

Evaluación de los Datos: Debemos determinar qué tipo de datos son de interés para la investigación (buzones de correo, ofimática, bases de datos...). Para ello, el primer paso es recuperar los datos que pudiera haber borrado el usuario con el fin de ocultar información. Como se puede ver en la Figura 12 es interesante determinar si se han instalado programas de borrado seguro dado que ello nos indica qué tipo de usuario es el propietario del dispositivo y si ha intentado ocultar información a conciencia. En tal caso, se deberá realizar una recuperación con métodos más exhaustivos a la recuperación automática.

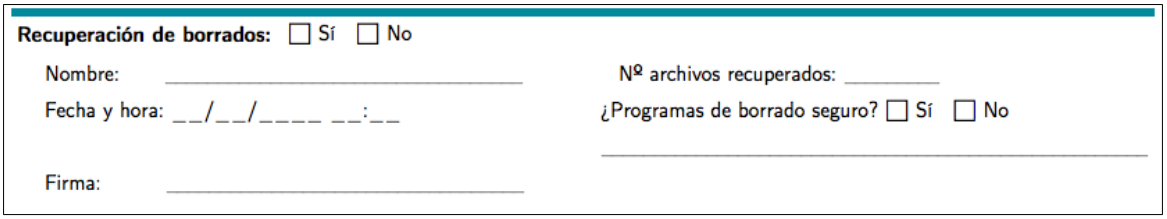


Figura 11: Documentación de la recuperación de archivos borrados.

En el desarrollo de esta guía no se va a entrar a detallar la recuperación no automatizada, pero es necesario recalcar que existen métodos más minuciosos para recuperar información parcial siempre que se tenga bien definidos los límites de la búsqueda. Un ejemplo es el lanzamiento de búsquedas en el espacio sin asignar del disco o aplicación de técnicas de *carving*.

Preparación de los Datos: Una vez se han recuperado todos los archivos posibles, se deben preparar los datos. Para ello realizaremos un análisis de firma con el fin de obtener un set de datos procesable. El análisis de firma se encarga de verificar que la extensión asignada al archivo corresponde con la realidad.

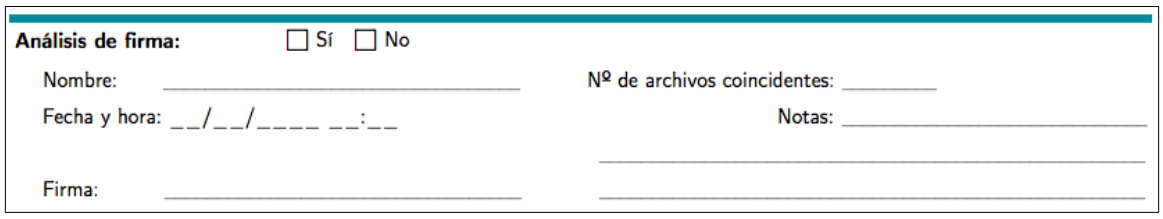


Figura 12: Documentación del resultado del Análisis de Firma.

En el ámbito general de las investigaciones empresariales se presupone que el usuario no tiene una motivación para ocultar información dentro del dispositivo a través de cambio de la firma de archivos, es por ello, que serán útiles para la investigación todos aquellos archivos que resulten tener firma coincidente.

Tras obtener un set de archivos válidos, seguiremos preparando los datos hasta obtener un set homogéneo y de primer nivel. Esto quiere decir, extraer la información de posibles contenedores hasta tener todos los archivos al mismo nivel. Los buzones de correo, backups y los archivos comprimidos son los contenedores más comunes.

A continuación, filtraremos los tipos de datos que son de interés para la investigación. Los tipos de archivos más comunes que pueden contener información relevante son: archivos de ofimática, buzones de correo y bases de datos. En la Figura 13 podemos ver cómo llevar un pequeño seguimiento de las características del set de datos acotado para la investigación. Con esta acción excluimos todos los archivos de programa y los propios del sistema operativo. Estos archivos son muy útiles para determinar el estado del dispositivo electrónico, como versión del sistema operativo, usuarios etc. El motivo por el que se excluyen de este procesamiento es porque son archivos que se deben analizar en función de las necesidades y no extraer de forma sistemática.

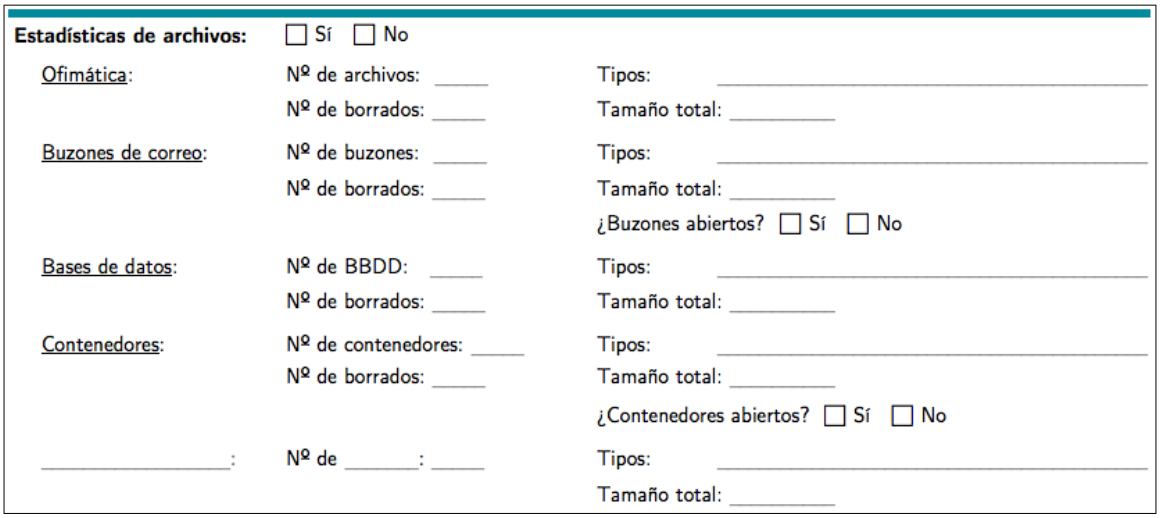


Figura 13: Estadísticas sobre la composición del set de datos.

Selección de los Datos: Una vez tenemos identificado los tipos de datos que pueden contener información relevante para la investigación debemos trabajar para dejar un set unificado

y preparado para revisar. Para ello, el siguiente paso será deduplicar los archivos para reducir el set de datos lo máximo posible. Hay tres formas de deduplicación: por evidencia, por custodian o global.

- **Deduplicación por evidencia:** La deduplicación por evidencia consiste en eliminar los duplicados dentro de una evidencia.
- **Deduplicación por custodian:** La deduplicación por *custodian* deduplica los archivos iterativamente de todas las evidencias de un mismo *custodian*. Esto es, si tenemos 3 evidencias, la primera se deduplicará sobre sí misma; la segunda sobre sí misma y la primera; y la tercera sobre sí misma, la primera y la segunda.
- **Deduplicación Global:** La deduplicación global consiste en deduplicar iterativamente sobre cada uno de los *custodians*. Sigue el mismo procedimiento que la deduplicación por *custodian*, pero ampliándola a varios *custodians*.

La deduplicación reduce considerablemente el set de datos dado que normalmente existen varias copias de un mismo archivo.

Además, se pueden identificar los **"casi-duplicados"** para agrupar las distintas versiones de un documento y así facilitar la revisión tratando conjuntamente todos estos documentos. Esta acción es altamente recomendable realizarla cuando se trabaja con documentos adjuntos a correos electrónicos puesto que en repetidas ocasiones se envía un documento modificado varias veces para la revisión de los últimos cambios.

Si tratamos con archivos de correo podemos seguir reduciendo el set a través de la identificación de **"hilos de correo"**. Consiste en agrupar los hilos de mensajes reuniendo todas las respuestas y mostrándolos como un solo documento o conjunto de documentos a revisar. Al realizar la agrupación de hilos de correo excluimos del set de datos a analizar todos aquellos correos que no sean inclusivos[4].

También se pueden aplicar **filtros temporales** para reducir el set de datos al periodo en el que se centra la investigación. Este proceso se puede aplicar al principio del procesamiento, teniendo como desventaja no poder acceder a los datos descartados a posteriori sin tener que repetir todos los pasos anteriores. La ventaja de realizar esta acción al principio radica en reducir considerablemente el set de datos y con ello el tiempo de procesamiento en las siguientes fases. Cada caso se debe tratar y analizar independientemente.

Normalización: En función de software utilizado, en muchas ocasiones es necesario normalizar los datos que serán la entrada de un software específico de revisión. Pasar todos los buzones de correo al mismo tipo de archivo (pst, ost...) sería un ejemplo de normalización.

Exportación: Si la plataforma de revisión y análisis no es la misma de procesamiento será necesario exportar los datos.

[4] "Un correo inclusivo es aquel que contiene contenido exclusivo no incluido en ningún otro correo electrónico. Un correo sin respuestas o reenvíos es un correo inclusivo. El último correo de un hilo es correo inclusivo." Relativity: https://help.kcura.com/9.0/Content/Relativity/Analytics/Inclusive_emails.htm

Validación: Como en todas las fases, se debe validar el proceso, comprobando en todo momento que las acciones realizadas sobre los datos son las encaminadas a acotar la investigación. Esta fase debe de hacerse en cada paso verificando que los procesos se llevan a cabo correctamente y comprobando que los resultados son coherentes con lo esperado.

3.3.2. Revisión y Análisis

Las etapas más importantes de la Revisión y Análisis son: Desarrollo de la Estrategia de Revisión, Formación del Equipo de Revisión, Análisis de los Datos y Flujo del Trabajo. Las fases de revisión y análisis son a medida de los proyectos, esta guía pretende orientar y dar un punto de partida extenso pero no cubrir todas las necesidades de todos los tipos de proyectos, deben de diseñarse para cada caso. Ambas fases deben de diseñarse de forma dinámica para poder adaptarse a los cambios y rumbos que pueda tomar la investigación, realizando iterativamente las fases en función de las novedades encontradas.

Desarrollo de la Estrategia de Revisión: La primera acción en esta fase del proyecto debe ser determinar el **alcance** y el **tipo de evidencia** que se desea encontrar. Una vez determinado, se debe diseñar el conjunto de palabras clave (también llamadas ***keywords***) y los **filtros** a emplear para reducir el set de datos a revisar según el objetivo de la investigación. Estos filtros pueden ser iguales a los aplicados en la fase anterior o filtros más específicos. La razón de aplicarlos en esta fase radica en su probabilidad de cambio en el transcurso de la investigación. Las *keywords* se pueden construir a través de **expresiones regulares** para abarcar las posibles combinaciones de un término o expresión y se lanzan sobre todo el texto de los archivos, incluyendo los metadatos. Los filtros se realizan a través de los metadatos de los archivos, pudiendo filtrar por remitente, por fecha o por tipo de archivo entre otros.

También se puede acotar el set de documentos a través de búsquedas de concepto y contexto como las Redes Neuronales, Métodos Bayesianos o Indexación Semántica. Estos métodos pueden fallar mucho en relevancia pero pueden ser muy útiles en investigaciones en las que el set de documentos sea muy grande o los términos de la investigación no se tengan claros todavía.

Se debe determinar la importancia y cómo se revisarán los documentos Casi-Duplicados para agilizar la revisión al equipo de revisión. Es muy importante determinar también el **tipo de formato** que adoptarán los documentos para su revisión, pudiendo ser nativos o convertidos a un estándar.

Para todo ello, se deben seleccionar adecuadamente las herramientas y la plataforma de revisión en función de las necesidades. Una plataforma escogida adecuadamente reduce la revisión, disminuyendo el tiempo de revisión y aumentando la calidad de la misma.

Por último, se debe crear un calendario donde se especifiquen los recursos asignados y el estado de la revisión en todo momento. Es vital cumplir con el presupuesto y los plazos de tiempo puesto que, como se ha comentado al inicio de este documento, en muchas ocasiones se trabaja con evidencias que acabarán en informes periciales en tribunales que marcan una estricta fecha de fin.

Formación del Equipo de Revisión: Es esencial hacer ver al equipo de revisión el alcance de la revisión y que toda la información con la que se trabaja es privilegiada.

Para llevar una gestión de los documentos es necesario establecer la definición de **Documento Relevante** y cómo etiquetarlo correctamente. Algunos de los aspectos por los que se puede catalogar un documento relevante pueden ser:

- Contiene información sensible.
- Categorización por tipo de hecho.
- Relación con los actores implicados.
- Contiene información confidencial (número de cuenta, teléfono, seguridad social, direcciones...).
- Indeterminado por encontrarse en otro idioma.

Además, se debe dejar un comentario explicativo sobre por qué el documento es relevante para la investigación.

Una vez definidos estos aspectos y entendidos por el equipo de revisión, éste se debe dividir en diferentes equipos para abordar las fases de revisión. Normalmente en revisiones de magnitudes medias se establecen dos niveles de revisión:

- El **Nivel 1**, formado normalmente por profesionales en sus primeros años en el campo de la investigación, que se encargan de descartar todos aquellos documentos no relevantes para la investigación y marcar como relevantes aquellos de los que se sospeche que puedan aportar información a la investigación. Se puede asignar la información en función de su sensibilidad o por custodian, por ejemplo.
- El **Nivel 2** formado por especialistas y responsables de la investigación, revisan sólo aquellos documentos que los revisores de primer nivel han marcado como relevantes.

Ambos niveles de revisores deben dejar sus comentarios sobre los documentos para una mejor clasificación y entendimiento.

Análisis de los Datos y Flujo del Trabajo: La fase de revisión y análisis es una de las más costosas en relación de tiempo y recursos dado que dependiendo el volumen de datos puede llegar a abarcar la mayor parte de la investigación. De ello, la importancia de relizar un seguimiento y un análisis del trabajo realizado. Una forma de cuantificar el seguimiento es a través de las **métricas de productividad** descritas a continuación:

- Número de documentos revisados por cada revisor en número de horas/días.
- Número de documentos marcados, por cada opción definida, por cada revisor.
- Número de documentos revisados hasta el momento.
- Número de documentos por revisar.

- etc ...

Combinando el porcentaje de avance de revisión y el número de documentos marcados como relevantes podemos hallar la **eficacia** de las *keywords* y búsquedas realizadas.

$$Eficacia = \frac{Número\ de\ documentos\ sensibles}{Número\ de\ documentos\ resultantes\ de\ keywords\ y\ búsquedas} \quad (1)$$

Si queremos observar la eficacia de las búsquedas podemos desagregar los documentos relevantes por *keywords*, búsquedas o aquellos criterios que se hayan establecido.

Es necesario resaltar la importancia, de nuevo, de llevar un orden y documentar el proceso para garantizar la trazabilidad de los documentos.

3.4. Producción y Presentación

La producción y presentación consiste en recolectar las evidencias finales y encajarlas en el marco de la investigación. Estas evidencias deben detallarse en un informe final dejando lo más claro posible la forma sobre cómo se han obtenido, y mostrando la trazabilidad de la investigación como se ha venido detallando en este proyecto. Para dar este soporte, el perito dispone de todos los formularios que se han ido diseñando en este proyecto.

Actualmente, dados los medios disponibles, estos informes se pueden producir digitalmente, permitiendo aportar las evidencias en formato nativo. Esto permite adjuntar la información tal y como se ha visto durante la investigación y exponer más información, como los metadatos o texto oculto, frente al formato de impresión.

En esta fase deben intervenir todos los especialistas en el campo de la investigación, los peritos en informática forense aportando todos los documentos relativos a cada evidencia y dando soporte técnico sobre su obtención y, por otro lado los técnicos en el área de la investigación, por ejemplo, financiera, deberán dar exponer el interés y el análisis realizado de la evidencia o conjunto de evidencias finales. Como en casi todas las fases, una vez más, la producción y presentación depende del caso concreto y debe realizarse a medida de la investigación. Así como garantizar la calidad y la rigurosidad de los procesos llevados a cabo deben estar entre los pilares de un proyecto eDiscovery.

4. Desarrollo de un caso práctico

En el marco de este proyecto, para visualizar mejor la metodología descrita, se va a realizar un caso de investigación práctico que desarrollará todas las fases de la guía eDiscovery de la primera sección de este documento. En el ámbito de la informática forense nos encontramos que no todos los proyectos son iguales y por ello se hace necesario adaptar los procedimientos a cada caso concreto, no obstante, todos comparten las mismas fases.

Se ha escogido el caso Enron, conocido como el mayor caso de fraude empresarial antes del 2001. Los rumores de pago de sobornos y tráfico de influencias fueron los detonantes de la investigación. Se ha escogido este caso por ser de dominio público y ofrecer el set de datos para su estudio.

El proyecto se va a dividir en dos fases. Acotada por los recursos disponibles para este proyecto, la primera fase describirá una hipotética acción de los investigadores forenses sobre la compañía, usando un ordenador personal para ilustrar cómo sería el procedimiento de extracción de información del mismo. La segunda fase, con los recursos obtenidos de internet, consistirá en realizar el procesamiento, revisión y análisis de los buzones de correo de diversos custodians de la compañía ENRON [6].

4.1. Identificación

Identificar cuáles van a ser las fuentes de información que nos proporcionarán los datos para resolver nuestro caso es el primer paso de toda investigación. [5]

Como se ha visto en la sección 3, la primera fase consiste en obtener información sobre cómo se gestiona la información dentro de la empresa. Basándonos en el formulario diseñado al respecto [Formulario B.1], se ha averiguado que existe una **Política de Gestión de la Información** actualizada conocida por todos los empleados y supervisada por el Departamento de Tecnologías de la Información. Además, se ha obtenido un listado de los empleados, que potencialmente podrían tener constancia de la información objeto de esta investigación, con los siguientes datos:

- Identificador de *Custodian*
- Identificador de Empleado
- Nombre y Apellidos
- Departamento Cargo
- Correo electrónico
- Teléfono de contacto

Tras analizar la información, procedemos a entrevistar al responsable del **Departamento de Tecnologías de la Información** para averiguar más sobre cómo se gestiona la información

[5] Todos los datos de los formularios han sido inventados, solo los nombres de los custodians están sacados del dataset de Enron.

en la empresa. [Formulario B.2]. Por otro lado, vamos elaborando un calendario de entrevistas para llegar a todos los empleados facilitados anteriormente, la empresa nos ha instado a encargarnos de, junto con un responsable suyo, explicar el proceso por el que se le requieren los dispositivos al usuario.

También se han planificado las entrevistas añadiendo al listado anterior el día y hora de la entrevista, así como el responsable de la misma. Durante la entrevista, se debe acordar un día y una hora de devolución. Esta fecha debería ser mínimo 24 horas posterior a la recolección del dispositivo con la especificación expresa de que no se garantiza que se pueda devolver el dispositivo en ese tiempo. Además, se debe especificar a la empresa que no se garantiza el correcto estado del dispositivo tras la devolución. Existen factores que pueden prolongar este tiempo, como por ejemplo, falta de material físico para adquirir la información de un dispositivo no estándar, dispositivos demasiado grandes que tomen más tiempo del esperado o fallo al adquirir la información que suponga su readquisición.

En este punto, iríamos siguiendo el plan de entrevistas, dadas las magnitudes nos centraremos en **Daren Farmer**. Podemos ver la simulación de la entrevista en el Formulario B.3.

4.2. Preservación y Adquisición

Para poder ilustrar esta fase del proyecto se ha escogido un ordenador personal modificando las fechas de adquisición para hacerlas coincidir con el marco de la investigación de Enron.

El primer lugar se debe firmar la cadena de custodia del dispositivo, Figura 14, y con el dispositivo en nuestro poder procedemos a preparar el entorno de trabajo para adquirirlo.

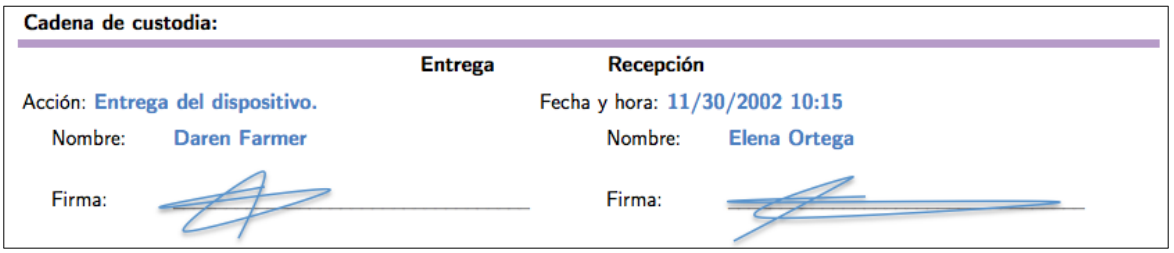
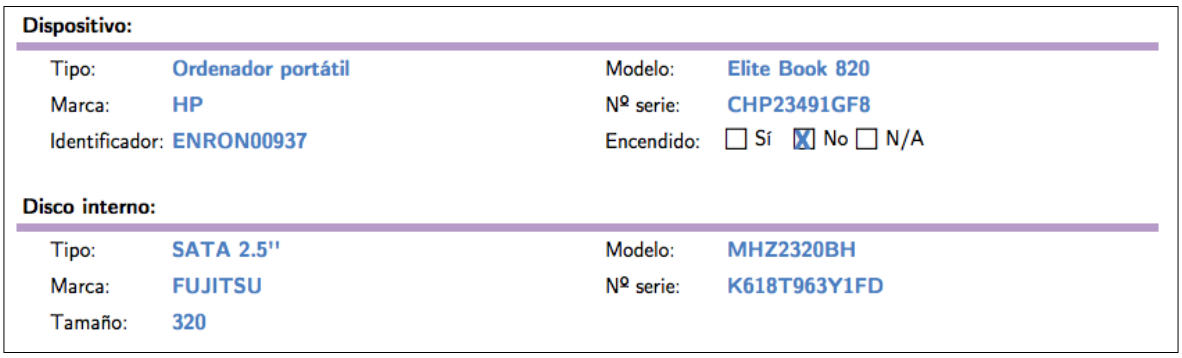


Figura 14: Cadena de custodia.

Como se ha comentado en la sección teórica, es importante identificar correctamente tanto el dispositivo como los elementos que pudiera contener dentro, en nuestro caso el ordenador y el disco duro interno. Se puede ver toda la información relativa a la Evidencia EPS00053 en la imagen 15. [Formulario B.4].

Una vez tenemos la evidencia identificada debemos preparar el disco destino que guardará la imagen. El disco destino debe ser de mayor tamaño que la evidencia; le llamaremos DESTINO023. Posteriormente, se realizará otra copia, en este caso comprimida, en el disco TRABAJO019.



Dispositivo:

Tipo:	Ordenador portátil	Modelo:	Elite Book 820
Marca:	HP	Nº serie:	CHP23491GF8
Identificador:	ENRON00937	Encendido:	☐ Sí ☒ No ☐ N/A

Disco interno:

Tipo:	SATA 2.5''	Modelo:	MHZ2320BH
Marca:	FUJITSU	Nº serie:	K618T963Y1FD
Tamaño:	320		

Figura 15: Información del dispositivo a adquirir.

4.2.1. Preparación de los discos de trabajo

Para preservar la información se va a proceder a cifrar tanto el disco destino como el de trabajo. En este proyecto se va a usar la herramienta VeraCrypt [5].

VeraCrypt permite cifrar tanto el disco completo o una partición, como la creación de un contenedor dinámico cifrado. En nuestro caso se realizará una partición cifrada.

Los discos cifrados son muy sensibles y si el programa deja de poder abrir la partición se perderá toda la información. Para minimizar este riesgo se deben guardar las cabeceras de la partición, preferiblemente dentro del mismo disco de destino utilizando una pequeña partición de 200MB sin cifrar. Si el disco se corrompiese, el programa VeraCrypt a través de la cabecera sería capaz de restaurarla e intentar descifrarlo.

Para nuestro ejemplo se usará el siguiente disco:

- Marca: WD
- Modelo: My passport 0820
- S/N: WXH1E649CYPL

Al igual que con las evidencias físicas de los custodians, se debe realizar un seguimiento de cada disco destino y de trabajo con el fin de saber quién lo creó, qué almacena y quién lo ha usado en cada momento. En el Formulario B.5 podemos ver la cadena de custodia de nuestro disco DESTINO0023.

4.2.2. Adquisición con EnCase

En este ejemplo práctico, se ha decidido trabajar con EnCase[7] como herramienta de adquisición. EnCase es un software de reconocido prestigio utilizado frecuentemente por empresas que se dedican al ámbito de la informática forense.

Una vez tenemos preparado el disco que va a almacenar nuestra evidencia debemos preparar la información a adquirir. En este caso, nos encontramos con un ordenador portátil apagado. En primer lugar, debemos extraer la batería y proceder a acceder al disco duro interno retirando

los tornillos necesarios.

Una vez hemos accedido al disco, lo desconectamos del ordenador y rellenamos los datos del disco interno en el Formulario B.4, Figura 15). Posteriormente lo conectamos a un bloqueador de escritura, Figura 16 y éste a nuestro equipo de trabajo.

Figura 16: Bloqueador de escritura Tableau.

Para comenzar con EnCase será necesario la creación de un nuevo caso y añadir nuestro disco duro extraído del ordenador del custodian al mismo. Posteriormente, seleccionando la opción de adquirir, procederemos a configurar la adquisición. Los parámetros más destacados son:

- **Nombre de la evidencia.** No debe contener datos que puedan identificarla con el Custodian para preservar lo más posible la protección de datos. Para aportar más información que el propio identificador, en el caso de ejemplo de la figura 17 se ha especificado el tipo de dispositivo y el tipo de adquisición.
- **Número de la evidencia.**
- **Notas.** Las notas deben contener toda la información posible para identificar a qué dispositivo pertenece la evidencia. En este caso, sí es aconsejable dejar reflejado el *custodian* al que pertenece, así como los datos del dispositivo y del disco interno.
- **Tamaño de los segmentos de la evidencia.** El tamaño de los segmentos de la evidencia nos permite crear segmentos de 1GB y así ver más fácilmente cuánto ocupa la evidencia.
- **Tipo de compresión.** EnCase permite comprimir la imagen. Ofrece dos niveles de compresión.
- **Hashes.** Obtención los hashes MD5 y SHA1.
- **Ruta destino.**

En la figura 17 podemos ver un ejemplo de la configuración para nuestro caso.

Una vez realizada la adquisición, deberemos esperar a que ésta se verifique correctamente. Esto nos asegura que se ha realizado correctamente y que la información es consistente y cumplirá el principio de confiabilidad. En la Tabla 1 podemos observar que la imagen se ha verificado correctamente.

En el anexo C podemos ver el reporte completo ofrecido por EnCase. Con la información obtenida del reporte continuamos rellenando la información solicitada en el Formulario B.4.

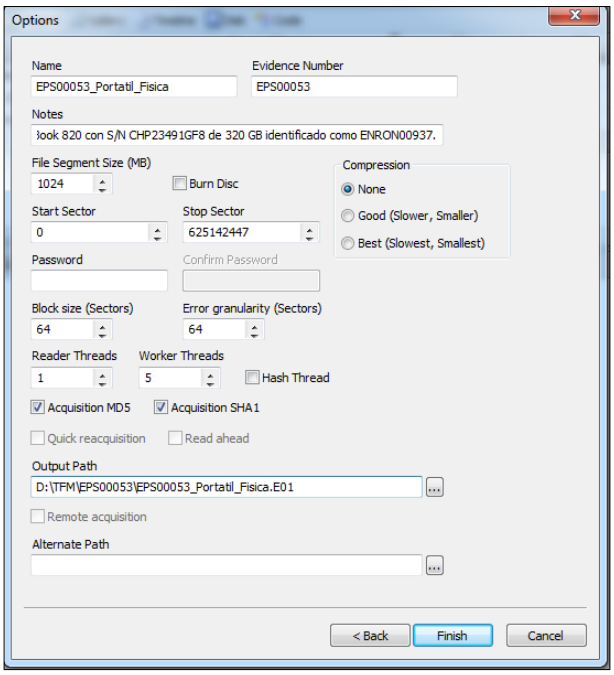


Figura 17: Encase: Configuración de la adquisición.

File Integrity	Completely Verified, 0 Errors
Acquisition MD5	b1de8f70f14dd6c296e731b99868edac
Acquisition MD5	b1de8f70f14dd6c296e731b99868edac
Acquisition SHA1	90892ba43fa8fcaecbfaed3573d77ee3b6aca2e4
Verification SHA1	90892ba43fa8fcaecbfaed3573d77ee3b6aca2e4

Tabla 1: Información parcial del reporte proporcionado por EnCase. [Anexo C].

Una vez finalizada la adquisición y documentada correctamente debemos devolver a quien pertenezca, o se haya acordado, el dispositivo, rellenando de nuevo la cadena de custodia. Figura 18.

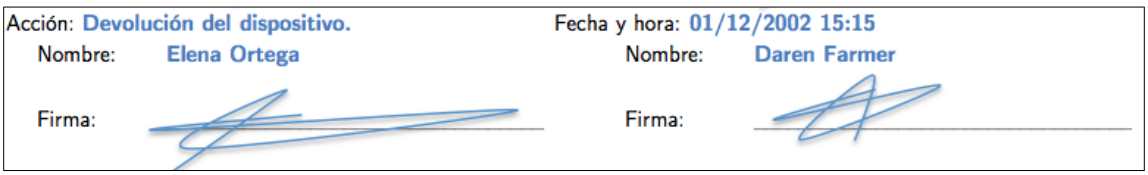


Figura 18: CoC: Devolución del dispositivo.

4.3. Procesamiento, Revisión y Análisis

4.3.1. Procesamiento

El procesamiento se va a realizar con dos herramientas: EnCase y Relativity. La primera fase del procesamiento se va a realizar con EnCase sobre la adquisición realizada en la sección anterior. Posteriormente se dejará de lado esta evidencia y se trabajará con el *dataset* de Enron y Relativity.

Evaluación de los datos: Realizaremos en primer lugar la recuperación de borrados mediante la opción *Recover Folders* de EnCase. Como se puede ver en la Figura 19 se han podido recuperar automáticamente 986.596 archivos de la imagen. En la figura 20 podemos ver cómo sería el seguimiento a dejar en el formulario B.6.

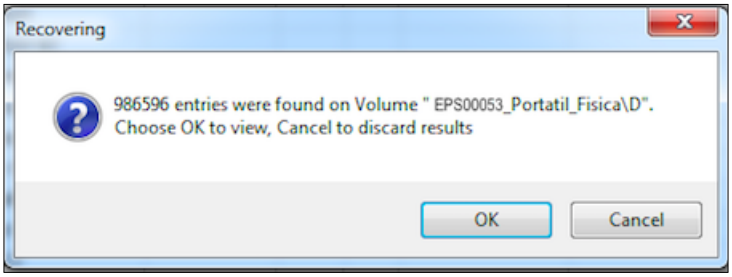


Figura 19: Recuperación de borrados.

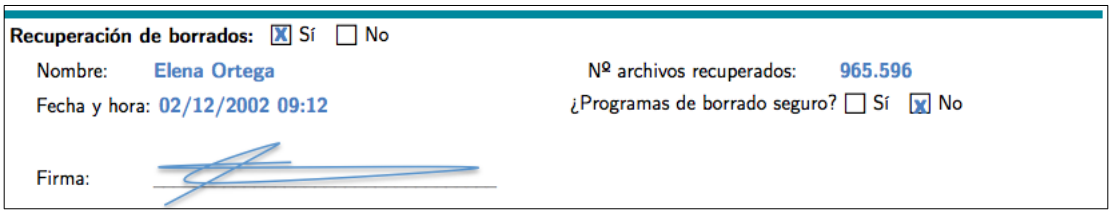


Figura 20: Recuperación de borrados.

Preparación de los datos: Una vez hemos recuperado los archivos, procedemos a preparar los datos realizando un análisis de firma. Desde EnCase podemos realizarlo mediante la opción *Verify file signatures*. Figura 21.

EnCase nos ofrece un pequeño reporte al acabar el proceso que utilizaremos para rellenar nuestro formulario B.6. Figura 22. Gracias a las notas podemos adecuar la información que nos aporta EnCase al formulario genérico diseñado en la primera fase de este proyecto.

Continuaremos preparando los datos aplicando filtros para seleccionar los archivos que resultan de interés para nuestra investigación. Para ello se han seleccionado las siguientes firmas de ofimática y correo electrónico. Tabla 2. Llegados a este punto, exportaremos los datos para importarlos en la plataforma Relativity.

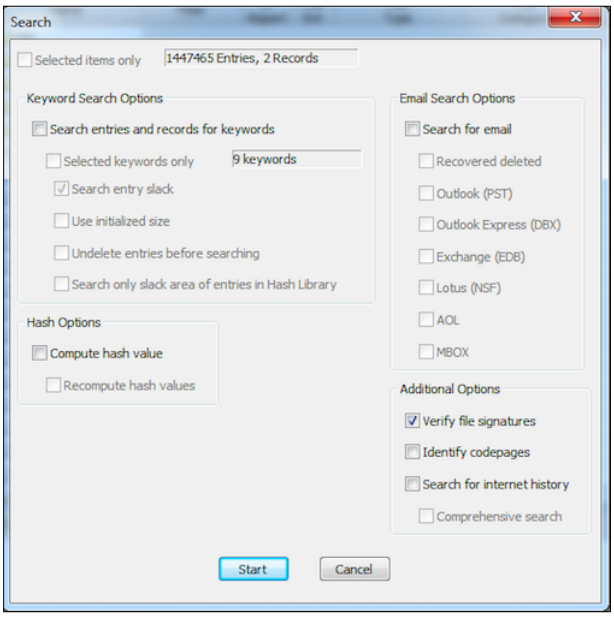


Figura 21: Análisis de firma.

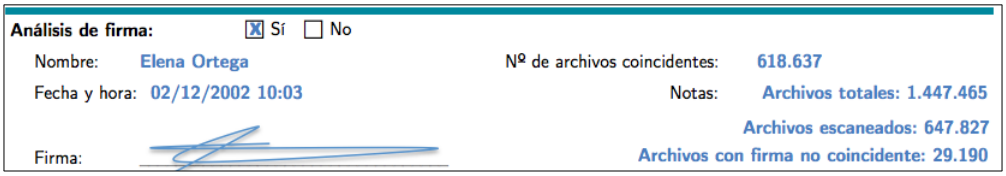


Figura 22: Formulario: Análisis de firma.

Firmas EnCase: Ofimática y correo
Compound Document File
UTF-8 Document File
XML Document
ZIP Compressed
Adobe PDF
HyperText Markup Language
Microsoft Access file
Rich Text Format
Gnu Zip Compressed
MS Compressed
* Compound Document File (MSG)
* Outlook Pst File
* Generic Email Message

Tabla 2: Firmas de ofimática y correo electrónico.

Llegados a este punto, se va a cambiar el set de datos al proporcionado por Enron y se va a ampliar el número de custodians a procesar. Los datos se han cogido de la red[6] puesto que son de dominio público, pero en un caso real alimentaríamos la plataforma con los exportados de EnCase. En concreto se va a trabajar con 130 custodians que se agruparán en cinco sets de procesamiento. El ejemplo anterior ilustra qué pasos deberían de darse por cada custodian para extraer la información. Relativity[8] es una de las plataformas más utilizadas y potentes actualmente para tratar masivamente datos en investigaciones.

El primer paso es cargar los datos en la plataforma. Para ello es necesario crear un Perfil de Procesamiento que configure qué tipo de datos vamos a cargar y qué tipo de parámetros queremos establecer. En la figura 23 podemos ver que el reconocimiento de texto se ha configurado en inglés, esto nos permite reconocer el texto en archivos en los que la extracción de texto no es automática. Además se ha configurado la extracción de todos aquellos documentos contenidos en otros, como por ejemplo los adjuntos. Respecto a la deduplicación, se ha escogido una deduplicación global para reducir al máximo la redundancia de los ficheros.

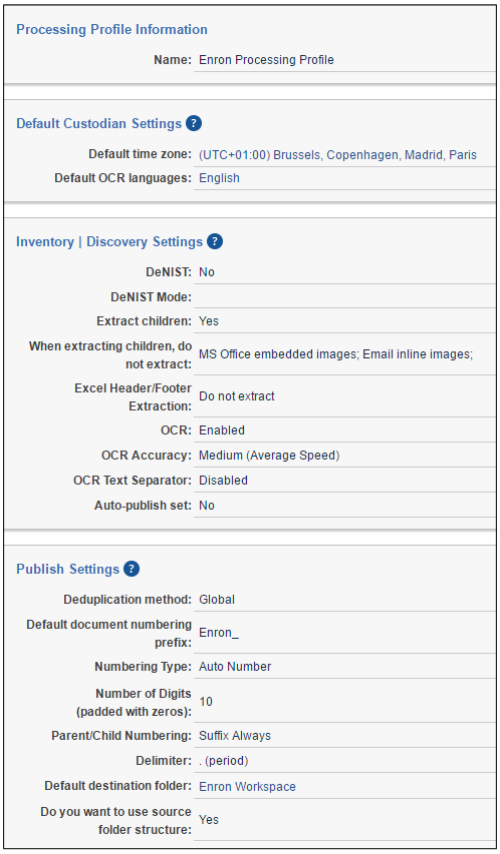


Figura 23: Relativity: Configuración del perfil de procesamiento.

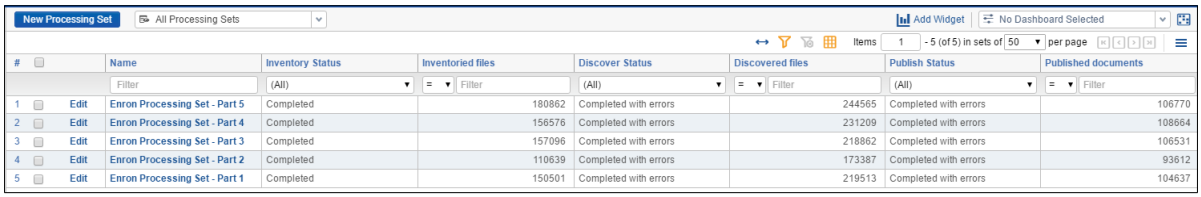
[6]Enron dataset: https://www.cs.cmu.edu/ ./enron/

En la figura 24 podemos ver las fases que lleva a cabo Relativity por cada set para cargar los datos y procesarlos. En primer lugar realiza un inventario de los ficheros, "*Inventory*", posteriormente descubre aquellos archivos contenidos en otros, por ejemplo los adjuntos de un correo, "*Discover*". Finalmente, publica los archivos finales con los que se va a trabajar, después de realizarles la deduplicación global como se ha especificado anteriormente,"*Publish*".

Figura 24: Estado del set de procesamiento.

Como se puede ver en la figura 24 existen errores que se pueden corregir. Estos errores se producen mayormente por archivos corruptos, protegidos por contraseña o imposibles de extraer los adjuntos a un correo. En este proyecto no se va a entrar a cómo tratarlos, pero se pueden corregir con distintas técnicas proporcionadas por Relativity, por ejemplo, una lista de posibles contraseñas para aplicar automáticamente a los archivos.

En la figura 25 podemos ver el total de sets procesados y el resultado, en número de documentos, de sus fases.


New Processing Set | All Processing Sets | Add Widget | No Dashboard Selected

Items 1 - 5 (of 5) in sets of 50 per page

#		Name	Inventory Status	Inventoried files	Discover Status	Discovered files	Publish Status	Published documents
		Filter	(All)	Filter	(All)	Filter	(All)	Filter
1	Edit	Enron Processing Set - Part 5	Completed	180862	Completed with errors	244565	Completed with errors	106770
2	Edit	Enron Processing Set - Part 4	Completed	156576	Completed with errors	231209	Completed with errors	108664
3	Edit	Enron Processing Set - Part 3	Completed	157096	Completed with errors	218862	Completed with errors	106531
4	Edit	Enron Processing Set - Part 2	Completed	110639	Completed with errors	173387	Completed with errors	93612
5	Edit	Enron Processing Set - Part 1	Completed	150501	Completed with errors	219513	Completed with errors	104637

Figura 25: Relativity: Total de sets procesados.

Tras cargar todos los documentos procedemos a configurar la agrupación de hilos de correo y su reducción asociada quedándonos solo con los correos inclusivos. En la figura 26 podemos ver la configuración de las cabeceras de los emails y de los metadatos. En la sección de opciones avanzadas se han definido las palabras que queremos que se excluyan del análisis, muy importante para posteriormente aplicar búsquedas por palabras clave.

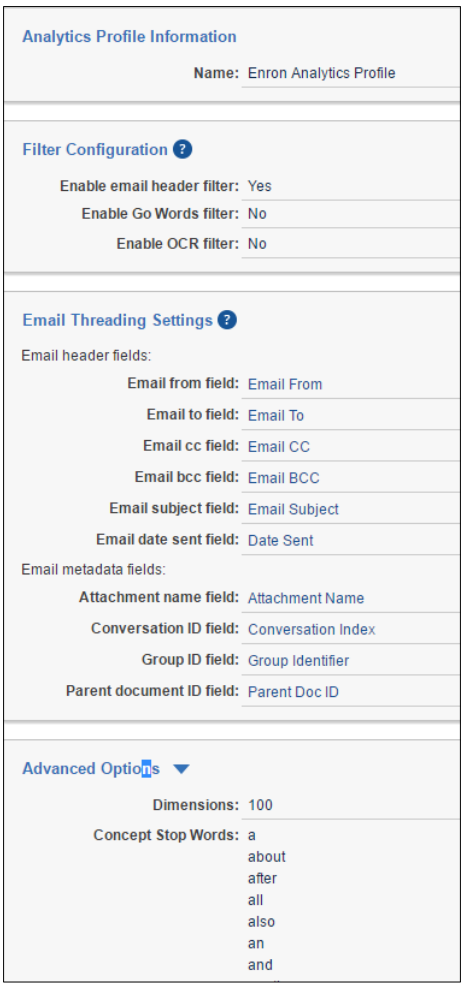


Figura 26: Relativity: Analytics Profile - Email Threading.

Para poder lanzar el proceso "*Email Threading*" debemos preparar el conjunto de datos siguiendo los pasos enunciados a continuación:

- Escoger solo aquellos archivos de tipo correo electrónico.
- Lanzar el proceso de identificación de los emails raíz denominados *Parent Emails*.
- Añadir la familia de estos documentos raíz.

Una vez tenemos el set preparado, que en nuestro caso se llama *A.005 Emails and Families*, configuramos el conjunto de análisis estructurado utilizando nuestro set y el perfil creado en el paso anterior. Figura 27.

Seguidamente, esta vez sobre el total de los documentos, preparamos otro conjunto de análisis estructurado para identificar todos aquellos documentos que se parecen en un 90 %. Esto nos permite agrupar las versiones de un mismo documento y facilitar la revisión. En la figura 28 podemos ver la configuración.

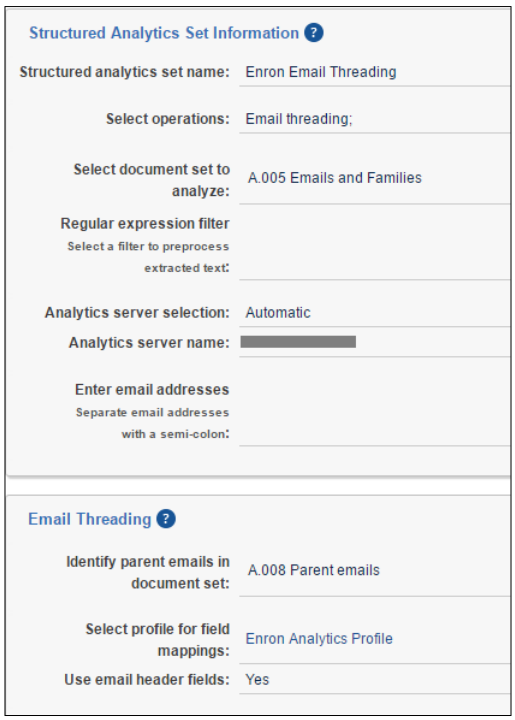


Figura 27: Relativity: *Structured Analytics Set - Email Threading.*

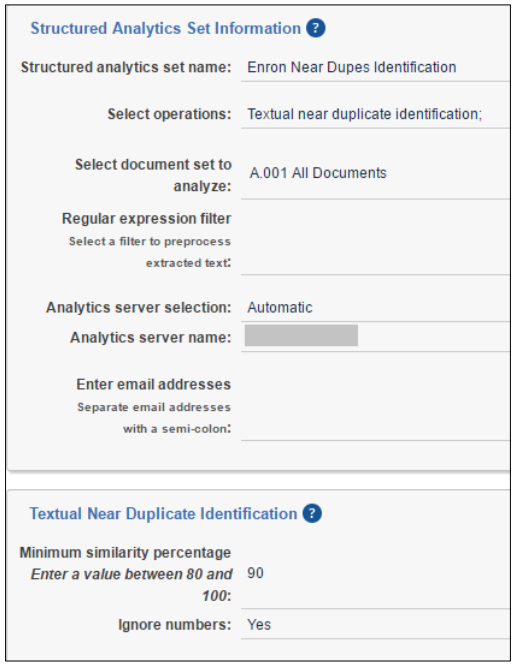


Figura 28: Relativity: *Structured Analytics Set - Near Duplicates.*

En este punto tenemos reducido al máximo el set de documentos e identificados los hilos de correo y agrupadas las versiones de un mismo documento. En este caso de ejemplo podemos dar por concluido el procesamiento y pasar a la fase de revisión.

4.3.2. Revisión y Análisis

Desarrollo de la estrategia de revisión: El alcance de esta investigación se delimita en la búsqueda, sobre el correo electrónico, de indicios que pongan de manifiesto que se conocían los delitos cometidos sobre la contabilidad de la empresa. Para ello se van a lanzar *keywords*, realizadas a través de expresiones regulares, para filtrar lo máximo posible los documentos y focalizar la investigación en aquellos que puedan tener conversaciones de más interés.

Para ello, el primer paso es construir un índice para poder realizar búsquedas eficientes sobre los documentos. Necesitamos escoger aquellos documentos en los cuales se pueda reconocer el texto y extraer. Como se ve en la figura 29 se ha escogido el set de documentos *A.002 All Documents - Extracted Text* para poder indexar el total de los documentos. Esto nos permite poder realizar búsquedas en el total de documentos y no sólo en el set reducido, así no habría que rehacer el índice si se necesitara otro set más ampliado. En este caso de ejemplo se ha escogido el tipo de índice dtSearch[9], basado en métodos de búsqueda tradicionales.

Con los documentos indexados podemos comenzar a definir los criterios de búsqueda. Este paso es muy específico de cada investigación, en este ejemplo se han elegido las siguientes palabras clave (*keywords*) para comenzar la investigación. Este proceso se puede repetir tantas veces sea necesario, en función de las necesidades que puedan surgir tras hallar nuevos descubrimientos.

Listado de palabras clave:

- *Energy*
- *Oil*
- *Futures*
- *Pipeline*
- *Compliance*
- *Pompano*
- *Jacoby*
- *Transmission*
- *Arbitrage*
- *Trad**
- *Risk*
- *Deal**
- *Product*
- *Gas*

En la figura 30 podemos ver los resultados en número de hits totales de las *keywords* anteriores y en el anexo D el reporte completo desglosado por cada *keyword*.

Una vez tenemos identificados los documentos que queremos revisar en primera instancia debemos preparar la plataforma para los revisores, los cuales se encargarán de analizar la información. Se van a establecer dos grupos, o *batches*, de documentos a partir de los obtenidos por las *keywords*, uno con los custodians prioritarios y otro con los no prioritarios. Figura 31. Cada *batch* será de máximo 250 documentos. Este tamaño debe ser estipulado en el plan de revisión para adecuarse a las necesidades y dinamizar lo máximo posible la revisión.

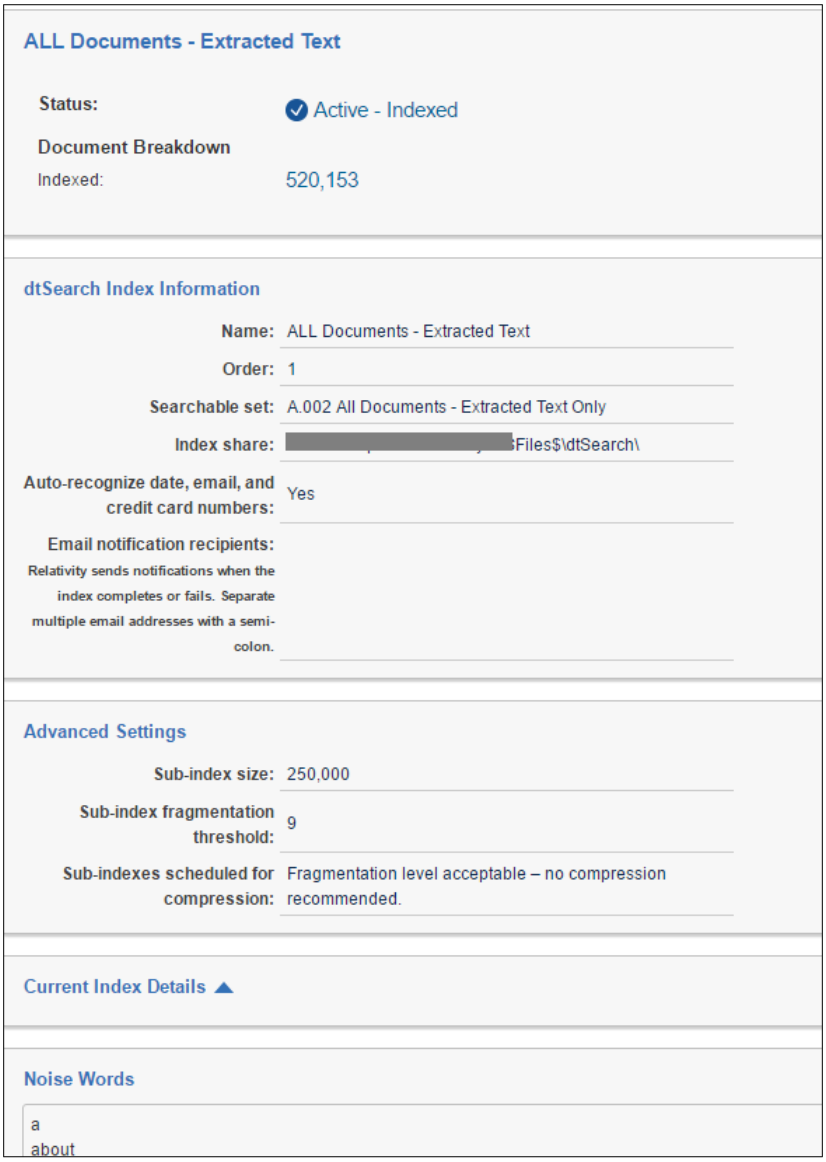


Figura 29: Relativity: Configuración del índice dtSearch.

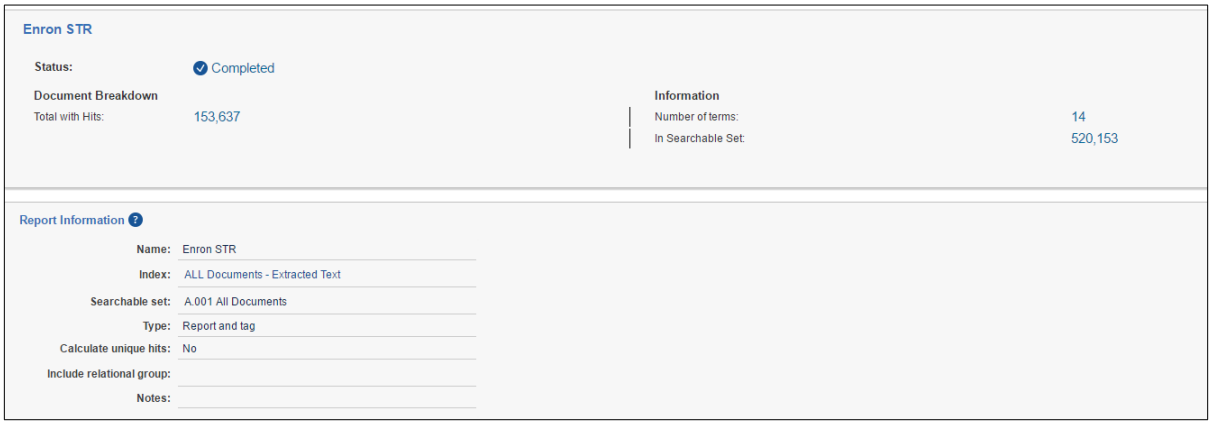


Figura 30: Relativity: Resultados de las *keywords*.

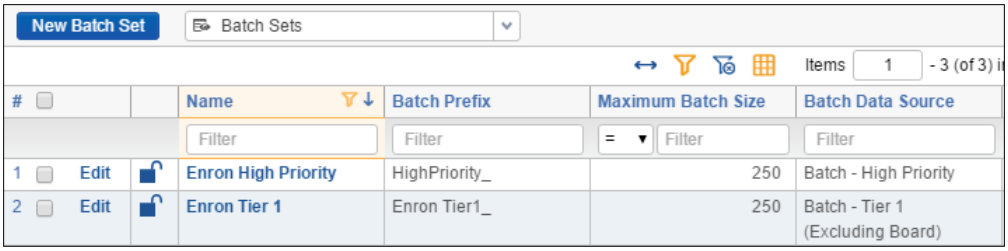


Figura 31: Relativity: Resultados de las *keywords*.

En la figura 32 podemos ver el total de sets de revisión creados. En este punto debemos establecer los grupos de revisión y los permisos que se les otorga.

#			Batch Set ↓	Batch	Assigned To	Batch Status	Batch Unit	Batch Size	Reviewed
			Filter	Filter	(All)	Filter	Filter	= Filter	= Filter
1		Edit	Enron Tier 1 Batch Set	Enron Tier1_00001		Completed		139	139
2		Edit	Enron Tier 1 Batch Set	Enron Tier1_00002		In Progress		69	4
3		Edit	Enron Tier 1 Batch Set	Enron Tier1_00003		In Progress		170	16
4		Edit	Enron Tier 1 Batch Set	Enron Tier1_00004		Completed		246	246
5		Edit	Enron Tier 1 Batch Set	Enron Tier1_00005		Completed		201	201
6		Edit	Enron Tier 1 Batch Set	Enron Tier1_00006		Completed		49	49
7		Edit	Enron Tier 1 Batch Set	Enron Tier1_00007		Completed		78	78
8		Edit	Enron Tier 1 Batch Set	Enron Tier1_00008		Completed		10	10
9		Edit	Enron Tier 1 Batch Set	Enron Tier1_00009		Completed		0	0

Figura 32: Relativity: Total de *batches*, asignación y progreso.

Se van a crear tres grupos de revisión. Los dos primeros equipos se encargarán de la primera revisión, separados por custodians de alta prioridad o no. El tercer equipo se encargará de la segunda revisión y de determinar las evidencias finales.

Para esta revisión se han especificado las siguientes etiquetas:

- Responsive (Relevante)
- Non-Responsive (No relevante)
- Needs Further Review (Categoría para que el revisor vuelva sobre él.)
- Technical Issue
- Foreing Language

  **Las dos últimas categorías pasarán a un paquete de documentos que se revisará en función de los recursos para resolver los problemas técnicos y de lenguaje.

Además, se pueden categorizar los documentos por su idioma principal:

- English
- French
- Other

Como ya se ha comentado, es muy importante disponer de un campo para dejar comentarios sobre el documento. En la Figura 33 se puede ver esta configuración.

En el caso del equipo de la segunda revisión, dispondrán de la configuración asignada por el grupo 1 y además un campo de revisado y un campo propio de comentarios. Figura 34.

Una vez finalizada la revisión se podrán exportar los documentos para su análisis exhaustivo y su posterior producción y presentación.

Los datos para realizar un análisis de los datos y del flujo de trabajo se pueden obtener de la Figura 32. Además, Relativity permite realizar tantas búsquedas como sean necesarias para obtener los datos para las métricas de revisión y relevancia de forma dinámica.

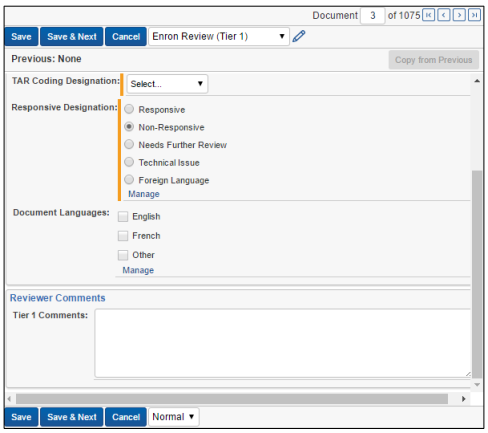


Figura 33: Relativity: Configuración del Nivel 1 de revisión.

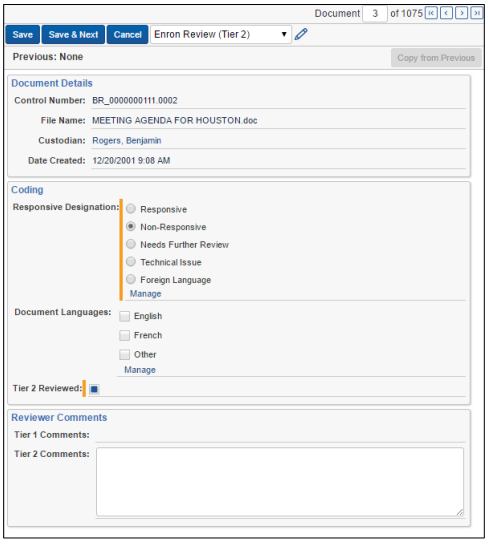


Figura 34: Relativity: Configuración del Nivel 2 de revisión.

4.4. Producción y Presentación

La producción y presentación se puede ofrecer en muchos formatos, en soporte digital o papel. El soporte digital ofrece la posibilidad de entregar las evidencias en en modo nativo o convertido a un mismo formato, por ejemplo, pdf. Ello dependerá de las normas legales por las que se rija el proyecto o el acuerdo tomado por el cliente.

Este proyecto no va a entrar a realizar un informe pericial. En este tipo de casos de varios investigados y diferentes fuentes de información se necesita un equipo numeroso y competente en el área de revisión para dilucidar las evidencias finales. No obstante, un ejemplo ilustrativo de cómo trazar como perito informático cada paso dado en la investigación, es este mismo proyecto. A través de los formularios y las acciones realizadas se puede detallar paso a paso cómo se ha obtenido cada evidencia que compone el caso.

Además, el informe deberá aportar información sobre la relevancia de las evidencias así como información que las soporte.

5. Conclusiones y trabajo futuro

Este proyecto cubre sólo una pequeña parte del mundo de las investigaciones forenses y se ha querido centraren una tipología concreta de investigación creando una pequeña guía de inicio. Ofrecer una buena base de formularios que permita asegurar las fases de Identificación, Preservación y Adquisición ha sido prioritario en este proyecto. Estas tres partes del diagrama EDRM son las más estándares a la hora de abordar un nuevo proyecto, siendo el resto más cambiantes en función de la investigación en cuestión. Como se ha podido ver en el caso práctico, la casuística de cada proyecto puede ser distinta y existen muchas vías de investigación y acciones a realizar para encontrar evidencias en una investigación. Dentro de un mismo tipo de investigación, los parámetros deben ajustarse a los recursos y las necesidades, creándose *ad hoc* para facilitar la investigación y centrarla lo máximo posible.

Este proyecto combina la normativa y el modelo estándar con la experiencia en casos reales, combinando ambas partes, teórica y práctica, para conseguir una guía que cubra el máximo tipo de proyectos posible dentro de una misma línea de investigaciones. La creación de los formularios ha sido meditada y analizada para diferentes proyectos con el fin de logar formularios que abarquen las máximas necesidades posibles. Respecto al caso práctico, ilustrar los pasos en un caso estándar de revisión de correo electrónico ha expuesto la cantidad de configuraciones posibles para un mismo proyecto en función del objetivo.

Como trabajo futuro, en este proyecto se han quedado sin cubrir un gran número de aspectos muy habituales en investigaciones forenses. Uno de ellos es el mundo de los artefactos de un sistema operativo que, en muchísimas ocasiones, son la clave para dilucidar el conjunto de acciones realizadas por el usuario en el dispositivo y no tanto la información que contiene, como se ha centrado este proyecto. También queda sin cubrir la extracción y análisis de memoria volátil, la cual ofrece mucha información sobre qué estaba haciendo el usuario justo en el momento en el que se requisa el dispositivo. Tampoco se cubre el análisis avanzado de acciones sobre el dispositivo, como la ya mencionada ocultación de información a través de firmas ficticias o la recuperación de información del espacio sin asignar del disco.

Respecto al análisis masivo de datos, las técnicas de análisis de conceptos mediante técnicas de aprendizaje automático darían para realizar muchas investigaciones al respecto que ayuden a acotar y guiar estas investigaciones.

Además, en el ejemplo práctico se ha hablado en todo momento de software privativo, muy adecuado a las necesidades, pero muy restrictivo a la hora de disponer de medios que permitan su costeo. Es por ello, la importancia de más investigación y desarrollo en herramientas de software libre que permitan la transferencia de conocimiento y se ajusten a las necesidades concretas de cada proyecto a través de desarrollo propio.

Referencias

[1] Introducción a la Informática Forense, Francisco Lázaro, 2013.

[2] ISO/IEC 27037:2012 Information technology – Security techniques – Guidelines for identification, collection, acquisition and preservation of digital evidence.

[3] *EDRM Identification Standards*, 22 de octubre del 2011, http://www.edrm.net/edrm-stages-standards/identification [Consulta: martes, 16 de febrero de 2016]

[4] *Cellebrite*, http://www.cellebrite.com [Consulta: lunes, 1 de agosto de 2016]

[5] *Veracrypt*, https://veracrypt.codeplex.com [Consulta: 6 de diciembre de 2016]

[6] Enron, https://es.wikipedia.org/wiki/Enron [Consulta: 7 de diciembre de 2016]

[7] EnCase, https://www.guidancesoftware.com/encase-forensic [Consulta 7 de diciembre de 2016]

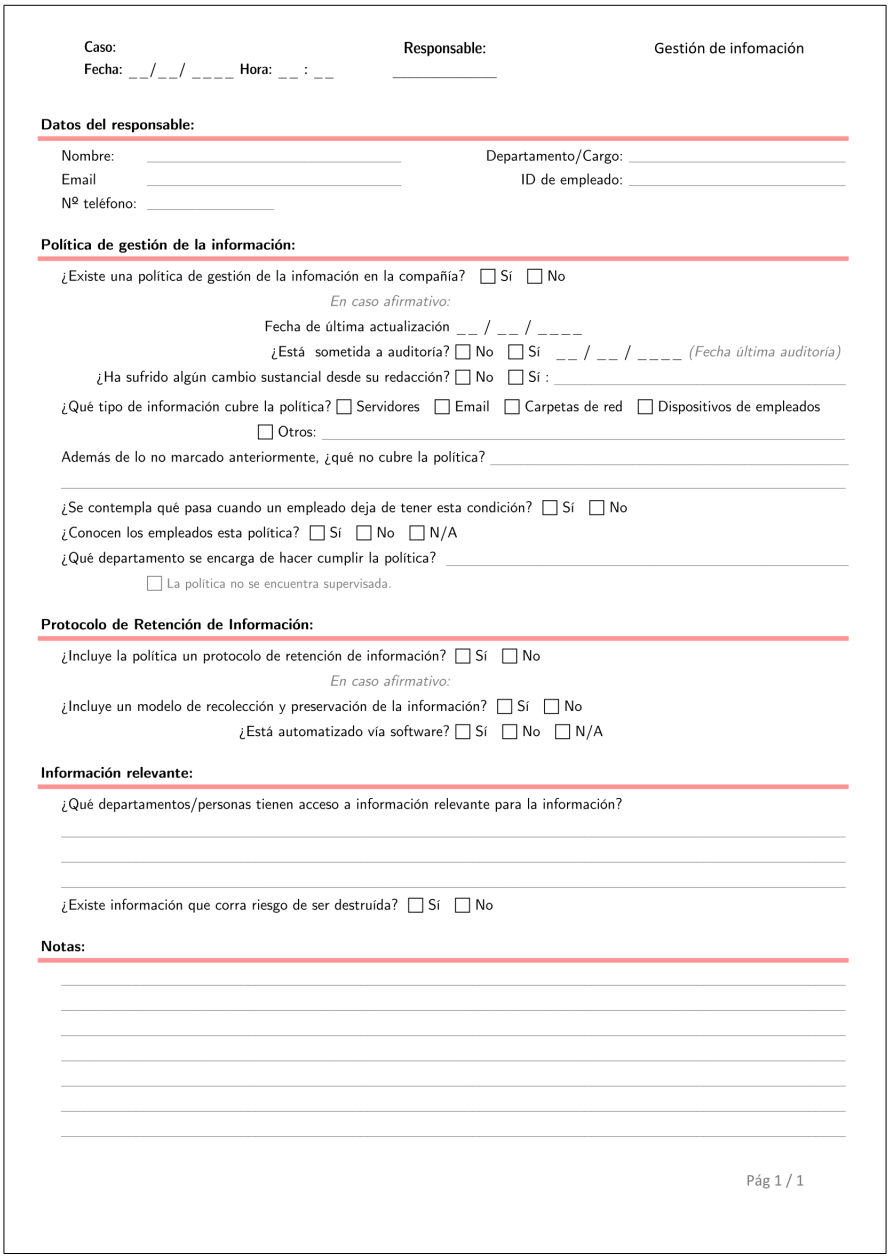
[8] Relativity, https://www.kcura.com/relativity/, [Consulta: 7 de diciembre de 2016]

[9] DtSearch - Relativity, https://www.kcura.com/relativity/Portals/0/Documents/8.0%20Documentation%20Help%20Site/Content/Features/dtSearch/dtSearch.htm, [Consulta: 7 de diciembre de 2016]

Anexos

A. Formularios

A.1. Gestión de Información



Caso:
Fecha: __/__/ ____ Hora: __ : __

Responsable: _____

Gestión de infomación

Datos del responsable:

Nombre: _____ Departamento/Cargo: _____
Email _____ ID de empleado: _____
Nº teléfono: _____

Política de gestión de la información:

¿Existe una política de gestión de la infomación en la compañía? ☐ Sí ☐ No

En caso afirmativo:

Fecha de última actualización __ / __ / ____
¿Está sometida a auditoría? ☐ No ☐ Sí __ / __ / ____ *(Fecha última auditoría)*
¿Ha sufrido algún cambio sustancial desde su redacción? ☐ No ☐ Sí : _____
¿Qué tipo de información cubre la política? ☐ Servidores ☐ Email ☐ Carpetas de red ☐ Dispositivos de empleados
☐ Otros: _____
Además de lo no marcado anteriormente, ¿qué no cubre la política? _____
¿Se contempla qué pasa cuando un empleado deja de tener esta condición? ☐ Sí ☐ No
¿Conocen los empleados esta política? ☐ Sí ☐ No ☐ N/A
¿Qué departamento se encarga de hacer cumplir la política? _____
☐ La política no se encuentra supervisada.

Protocolo de Retención de Información:

¿Incluye la política un protocolo de retención de información? ☐ Sí ☐ No

En caso afirmativo:

¿Incluye un modelo de recolección y preservación de la información? ☐ Sí ☐ No
¿Está automatizado vía software? ☐ Sí ☐ No ☐ N/A

Información relevante:

¿Qué departamentos/personas tienen acceso a información relevante para la información?

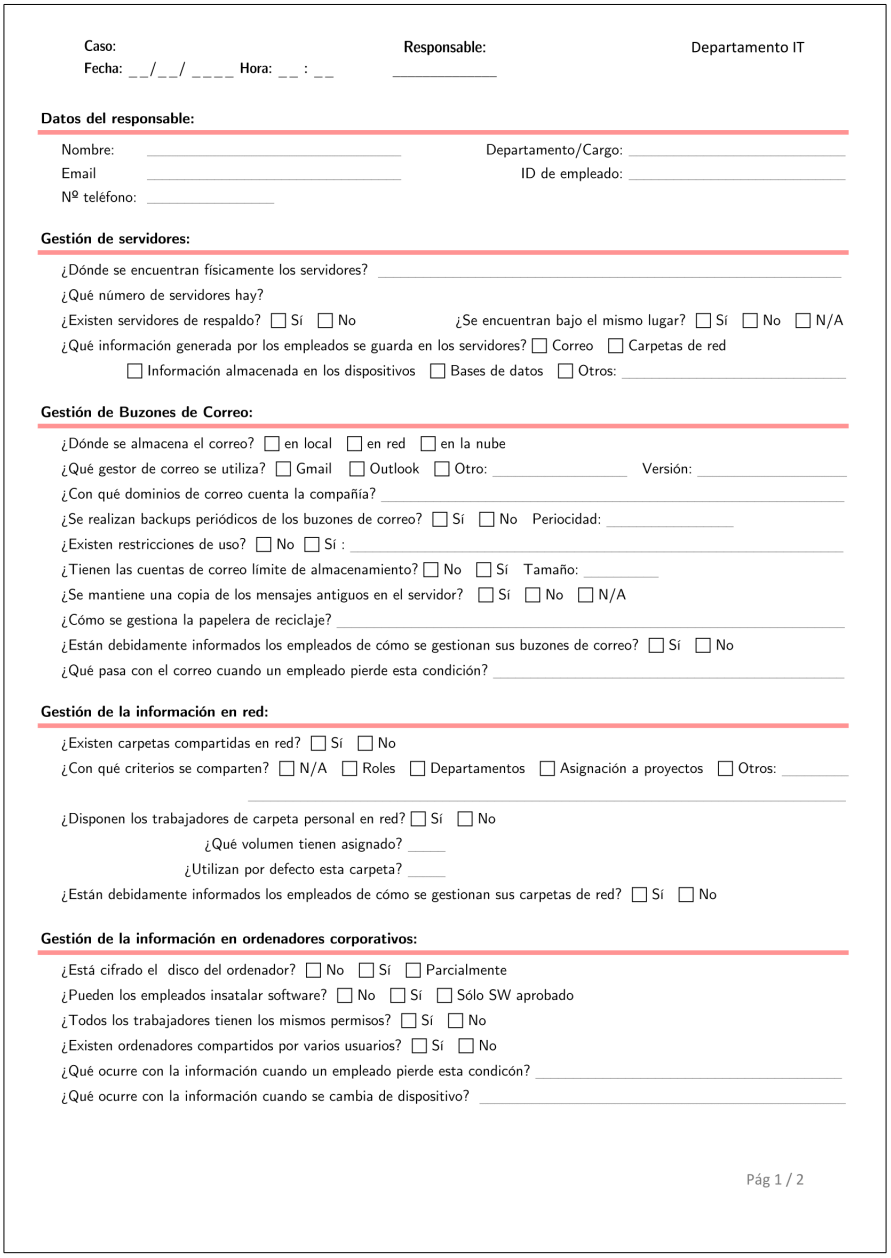
¿Existe información que corra riesgo de ser destruída? ☐ Sí ☐ No

Notas:

Pág 1 / 1

Figura A1: Formulario Gestión de Información.

A.2. Departamento IT



Caso: Responsable: Departamento IT

Fecha: __/__/ ____ Hora: __ : __

Datos del responsable:

Nombre: _____ Departamento/Cargo: _____

Email _____ ID de empleado: _____

Nº teléfono: _____

Gestión de servidores:

¿Dónde se encuentran fisicamente los servidores? _____

¿Qué número de servidores hay?

¿Existen servidores de respaldo? ☐ Sí ☐ No ¿Se encuentran bajo el mismo lugar? ☐ Sí ☐ No ☐ N/A

¿Qué información generada por los empleados se guarda en los servidores? ☐ Correo ☐ Carpetas de red

☐ Información almacenada en los dispositivos ☐ Bases de datos ☐ Otros: _____

Gestión de Buzones de Correo:

¿Dónde se almacena el correo? ☐ en local ☐ en red ☐ en la nube

¿Qué gestor de correo se utiliza? ☐ Gmail ☐ Outlook ☐ Otro: _____ Versión: _____

¿Con qué dominios de correo cuenta la compañía? _____

¿Se realizan backups periódicos de los buzones de correo? ☐ Sí ☐ No Periocidad: _____

¿Existen restricciones de uso? ☐ No ☐ Sí : _____

¿Tienen las cuentas de correo límite de almacenamiento? ☐ No ☐ Sí Tamaño: _____

¿Se mantiene una copia de los mensajes antiguos en el servidor? ☐ Sí ☐ No ☐ N/A

¿Cómo se gestiona la papelera de reciclaje? _____

¿Están debidamente informados los empleados de cómo se gestionan sus buzones de correo? ☐ Sí ☐ No

¿Qué pasa con el correo cuando un empleado pierde esta condición? _____

Gestión de la información en red:

¿Existen carpetas compartidas en red? ☐ Sí ☐ No

¿Con qué criterios se comparten? ☐ N/A ☐ Roles ☐ Departamentos ☐ Asignación a proyectos ☐ Otros: _____

_____

¿Disponen los trabajadores de carpeta personal en red? ☐ Sí ☐ No

¿Qué volumen tienen asignado? _____

¿Utilizan por defecto esta carpeta? _____

¿Están debidamente informados los empleados de cómo se gestionan sus carpetas de red? ☐ Sí ☐ No

Gestión de la información en ordenadores corporativos:

¿Está cifrado el disco del ordenador? ☐ No ☐ Sí ☐ Parcialmente

¿Pueden los empleados insatalar software? ☐ No ☐ Sí ☐ Sólo SW aprobado

¿Todos los trabajadores tienen los mismos permisos? ☐ Sí ☐ No

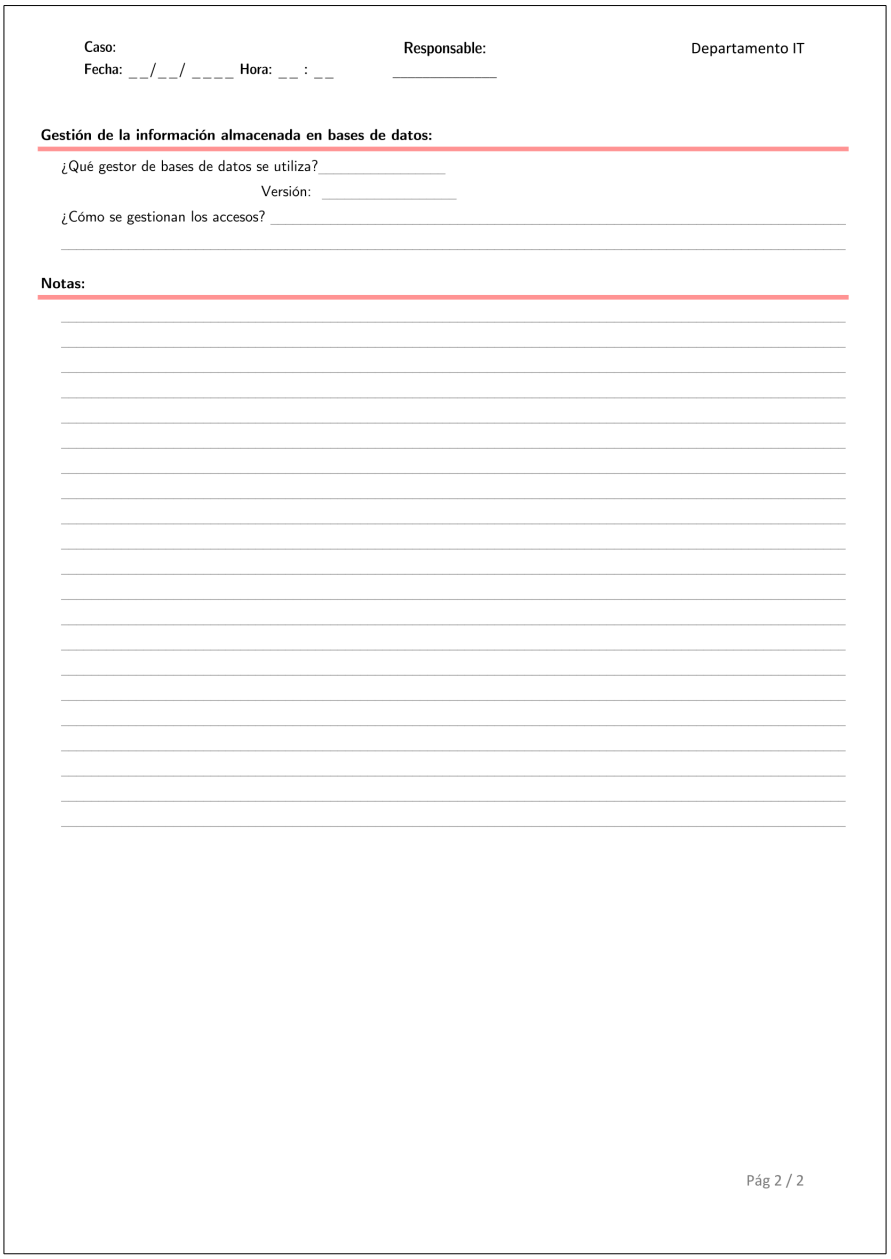
¿Existen ordenadores compartidos por varios usuarios? ☐ Sí ☐ No

¿Qué ocurre con la información cuando un empleado pierde esta condicón? _____

¿Qué ocurre con la información cuando se cambia de dispositivo? _____

Pág 1 / 2

Figura A2: Formulario Departamento IT (Anverso).


Caso: Responsable: Departamento IT
Fecha: __/__/ ____ Hora: __ : __ _____

Gestión de la información almacenada en bases de datos:

¿Qué gestor de bases de datos se utiliza?_____

Versión: _____

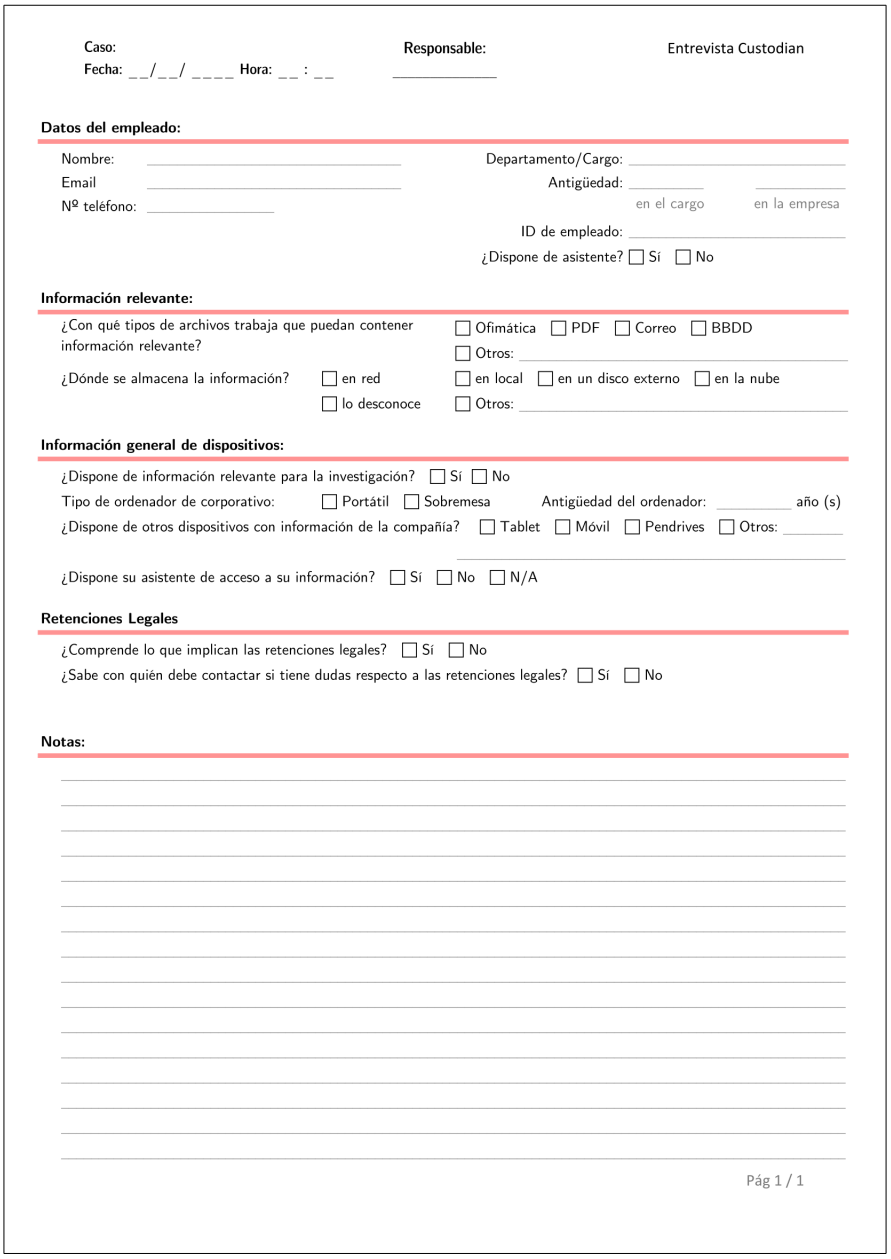
¿Cómo se gestionan los accesos? _____

Notas:

Pág 2 / 2

Figura A3: Formulario Departamento IT (Reverso).

A.3. Entrevista Custodian



Caso:
Responsable:
Entrevista Custodian
Fecha: __/__/ ____ Hora: __ : __

Datos del empleado:

Nombre:
Email
Nº teléfono:
Departamento/Cargo:
Antigüedad: en el cargo en la empresa
ID de empleado:
¿Dispone de asistente? ☐ Sí ☐ No

Información relevante:

¿Con qué tipos de archivos trabaja que puedan contener información relevante? ☐ Ofimática ☐ PDF ☐ Correo ☐ BBDD ☐ Otros:
¿Dónde se almacena la información? ☐ en red ☐ lo desconoce ☐ en local ☐ en un disco externo ☐ en la nube ☐ Otros:

Información general de dispositivos:

¿Dispone de información relevante para la investigación? ☐ Sí ☐ No
Tipo de ordenador de corporativo: ☐ Portátil ☐ Sobremesa Antigüedad del ordenador: año (s)
¿Dispone de otros dispositivos con información de la compañía? ☐ Tablet ☐ Móvil ☐ Pendrives ☐ Otros:
¿Dispone su asistente de acceso a su información? ☐ Sí ☐ No ☐ N/A

Retenciones Legales

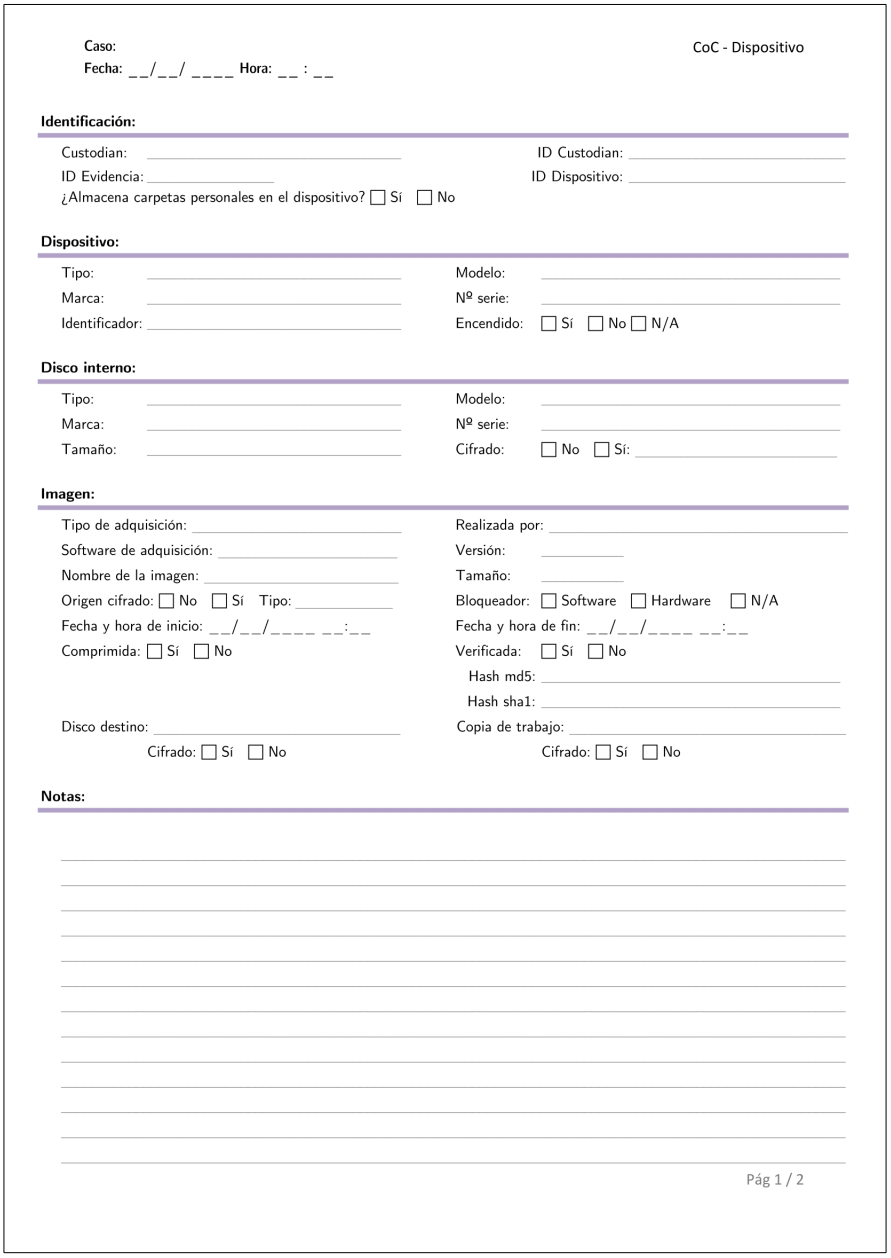
¿Comprende lo que implican las retenciones legales? ☐ Sí ☐ No
¿Sabe con quién debe contactar si tiene dudas respecto a las retenciones legales? ☐ Sí ☐ No

Notas:

Pág 1 / 1

Figura A4: Formulario Entrevista Custodian.

A.4. CoC - Dispositivo



Caso:

Fecha: __/__/ ____ Hora: __ : __

CoC - Dispositivo

Identificación:

Custodian: _____ ID Custodian: _____

ID Evidencia: _____ ID Dispositivo: _____

¿Almacena carpetas personales en el dispositivo? ☐ Sí ☐ No

Dispositivo:

Tipo: _____ Modelo: _____

Marca: _____ Nº serie: _____

Identificador: _____ Encendido: ☐ Sí ☐ No ☐ N/A

Disco interno:

Tipo: _____ Modelo: _____

Marca: _____ Nº serie: _____

Tamaño: _____ Cifrado: ☐ No ☐ Sí: _____

Imagen:

Tipo de adquisición: _____ Realizada por: _____

Software de adquisición: _____ Versión: _____

Nombre de la imagen: _____ Tamaño: _____

Origen cifrado: ☐ No ☐ Sí Tipo: _____ Bloqueador: ☐ Software ☐ Hardware ☐ N/A

Fecha y hora de inicio: __/__/____ __:__ Fecha y hora de fin: __/__/____ __:__

Comprimida: ☐ Sí ☐ No Verificada: ☐ Sí ☐ No

Hash md5: _____

Hash sha1: _____

Disco destino: _____ Copia de trabajo: _____

Cifrado: ☐ Sí ☐ No Cifrado: ☐ Sí ☐ No

Notas:

Pág 1 / 2

Figura A5: Formulario Cadena de Custodia - Dispositivo (Anverso).

Caso: CoC - Dispositivo

Fecha: __/__/ ____ **Hora:** __ : __

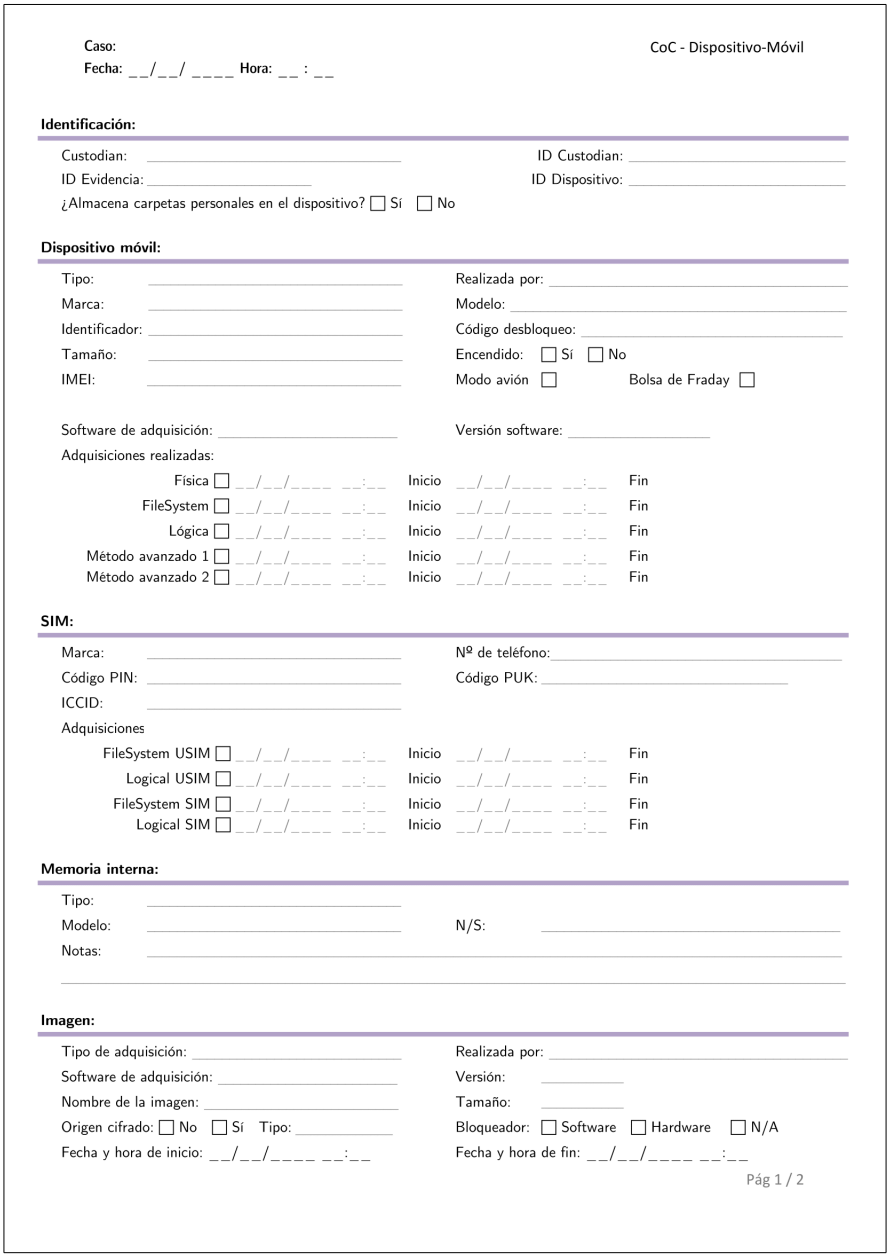
Cadena de custodia:

Acción: ______	Entrega	Recepción
	Nombre: ______	Nombre: ______
	Fecha y hora: __/__/____ __:__	Fecha y hora: __/__/____ __:__
	Firma: ______	Firma: ______
Acción: ______		
	Nombre: ______	Nombre: ______
	Fecha y hora: __/__/____ __:__	Fecha y hora: __/__/____ __:__
	Firma: ______	Firma: ______
Acción: ______		
	Nombre: ______	Nombre: ______
	Fecha y hora: __/__/____ __:__	Fecha y hora: __/__/____ __:__
	Firma: ______	Firma: ______
Acción: ______		
	Nombre: ______	Nombre: ______
	Fecha y hora: __/__/____ __:__	Fecha y hora: __/__/____ __:__
	Firma: ______	Firma: ______
Acción: ______		
	Nombre: ______	Nombre: ______
	Fecha y hora: __/__/____ __:__	Fecha y hora: __/__/____ __:__
	Firma: ______	Firma: ______
Acción: ______		
	Nombre: ______	Nombre: ______
	Fecha y hora: __/__/____ __:__	Fecha y hora: __/__/____ __:__
	Firma: ______	Firma: ______
Acción: ______		
	Nombre: ______	Nombre: ______
	Fecha y hora: __/__/____ __:__	Fecha y hora: __/__/____ __:__
	Firma: ______	Firma: ______
Acción: ______		
	Nombre: ______	Nombre: ______
	Fecha y hora: __/__/____ __:__	Fecha y hora: __/__/____ __:__
	Firma: ______	Firma: ______

Pág 2 / 2

Figura A6: Formulario Cadena de Custodia - Dispositivo (Reverso).

A.5. CoC - Dispositivo Móvil



Caso:

Fecha: __/__/ ____ Hora: __ : __

CoC - Dispositivo-Móvil

Identificación:

Custodian: _____ ID Custodian: _____

ID Evidencia: _____ ID Dispositivo: _____

¿Almacena carpetas personales en el dispositivo? ☐ Sí ☐ No

Dispositivo móvil:

Tipo: _____ Realizada por: _____

Marca: _____ Modelo: _____

Identificador: _____ Código desbloqueo: _____

Tamaño: _____ Encendido: ☐ Sí ☐ No

IMEI: _____ Modo avión ☐ Bolsa de Fraday ☐

Software de adquisición: _____ Versión software: _____

Adquisiciones realizadas:

Física ☐	__/__/____ __:__	Inicio	__/__/____ __:__	Fin
FileSystem ☐	__/__/____ __:__	Inicio	__/__/____ __:__	Fin
Lógica ☐	__/__/____ __:__	Inicio	__/__/____ __:__	Fin
Método avanzado 1 ☐	__/__/____ __:__	Inicio	__/__/____ __:__	Fin
Método avanzado 2 ☐	__/__/____ __:__	Inicio	__/__/____ __:__	Fin

SIM:

Marca: _____ Nº de teléfono: _____

Código PIN: _____ Código PUK: _____

ICCID: _____

Adquisiciones

FileSystem USIM ☐	__/__/____ __:__	Inicio	__/__/____ __:__	Fin
Logical USIM ☐	__/__/____ __:__	Inicio	__/__/____ __:__	Fin
FileSystem SIM ☐	__/__/____ __:__	Inicio	__/__/____ __:__	Fin
Logical SIM ☐	__/__/____ __:__	Inicio	__/__/____ __:__	Fin

Memoria interna:

Tipo: _____

Modelo: _____ N/S: _____

Notas: _____

Imagen:

Tipo de adquisición: _____ Realizada por: _____

Software de adquisición: _____ Versión: _____

Nombre de la imagen: _____ Tamaño: _____

Origen cifrado: ☐ No ☐ Sí Tipo: _____ Bloqueador: ☐ Software ☐ Hardware ☐ N/A

Fecha y hora de inicio: __/__/____ __:__ Fecha y hora de fin: __/__/____ __:__

Pág 1 / 2

Figura A7: Formulario Cadena de Custodia - Dispositivo Móvil (Anverso).

Caso: CoC - Dispositivo-Móvil

Fecha: __/__/ ____ **Hora:** __ : __

Comprimida: ☐ Sí ☐ No

Verificada: ☐ Sí ☐ No

Hash md5: ____________________

Hash sha1: ____________________

Disco destino: ____________________

Cifrado: ☐ Sí ☐ No

Copia de trabajo: ____________________

Cifrado: ☐ Sí ☐ No

Notas:

Cadena de custodia:

Acción: __________

Entrega	**Recepción**
Nombre: __________	Nombre: __________
Fecha y hora: __/__/____ __:__	Fecha y hora: __/__/____ __:__
Firma: __________	Firma: __________

Acción: __________

Entrega	**Recepción**
Nombre: __________	Nombre: __________
Fecha y hora: __/__/____ __:__	Fecha y hora: __/__/____ __:__
Firma: __________	Firma: __________

Acción: __________

Entrega	**Recepción**
Nombre: __________	Nombre: __________
Fecha y hora: __/__/____ __:__	Fecha y hora: __/__/____ __:__
Firma: __________	Firma: __________

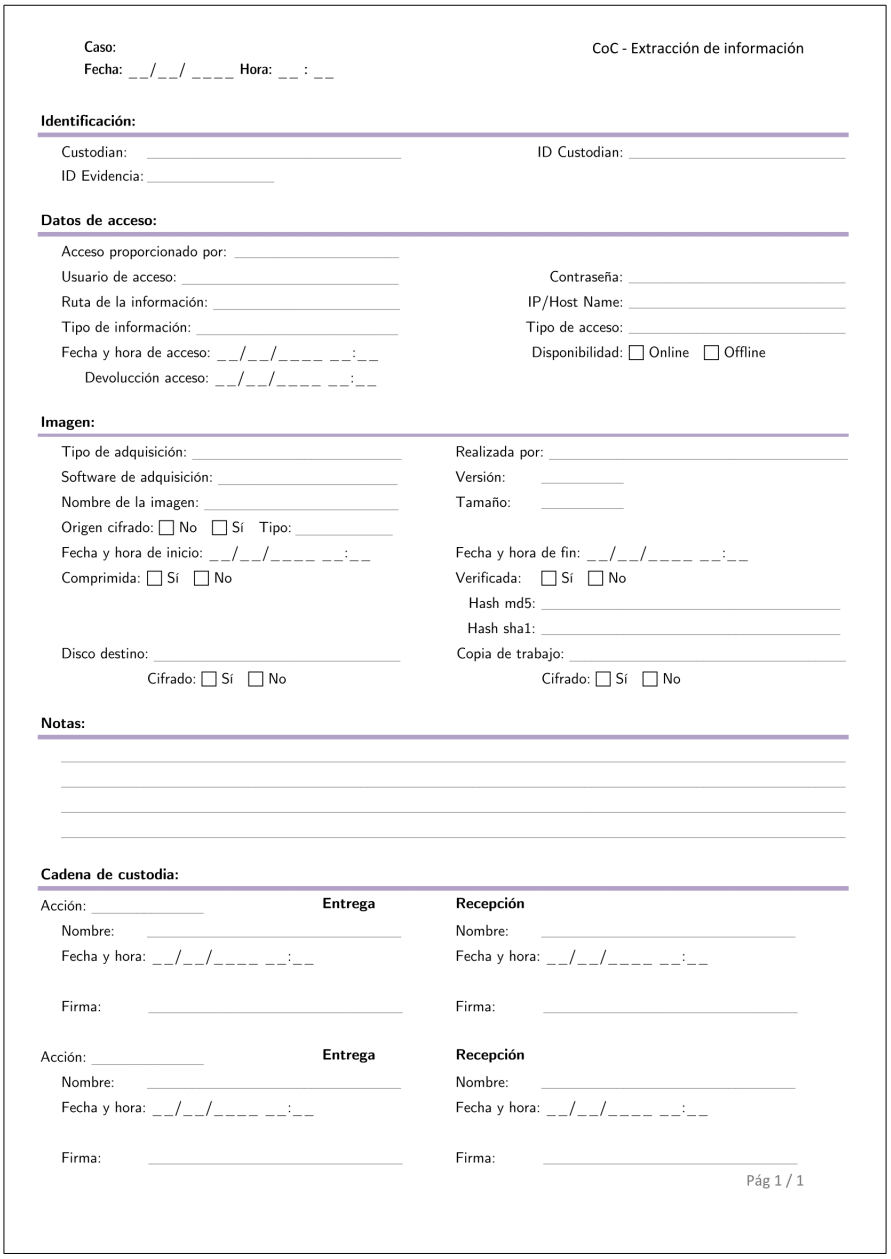
Acción: __________

Entrega	**Recepción**
Nombre: __________	Nombre: __________
Fecha y hora: __/__/____ __:__	Fecha y hora: __/__/____ __:__
Firma: __________	Firma: __________

Pág 2 / 2

Figura A8: Formulario Cadena de Custodia - Dispositivo Móvil (Reverso).

A.6. CoC - Extracción de Información



Caso:

CoC - Extracción de información

Fecha: __/__/ ____ Hora: __ : __

Identificación:

Custodian: _____ ID Custodian: _____

ID Evidencia: _____

Datos de acceso:

Acceso proporcionado por: _____

Usuario de acceso: _____ Contraseña: _____

Ruta de la información: _____ IP/Host Name: _____

Tipo de información: _____ Tipo de acceso: _____

Fecha y hora de acceso: __/__/____ __:__ Disponibilidad: ☐ Online ☐ Offline

Devolucción acceso: __/__/____ __:__

Imagen:

Tipo de adquisición: _____ Realizada por: _____

Software de adquisición: _____ Versión: _____

Nombre de la imagen: _____ Tamaño: _____

Origen cifrado: ☐ No ☐ Sí Tipo: _____

Fecha y hora de inicio: __/__/____ __:__ Fecha y hora de fin: __/__/____ __:__

Comprimida: ☐ Sí ☐ No Verificada: ☐ Sí ☐ No

Hash md5: _____

Hash sha1: _____

Disco destino: _____ Copia de trabajo: _____

Cifrado: ☐ Sí ☐ No Cifrado: ☐ Sí ☐ No

Notas:

_____

_____

_____

_____

Cadena de custodia:

Acción: _____ **Entrega** **Recepción**

Nombre: _____ Nombre: _____

Fecha y hora: __/__/____ __:__ Fecha y hora: __/__/____ __:__

Firma: _____ Firma: _____

Acción: _____ **Entrega** **Recepción**

Nombre: _____ Nombre: _____

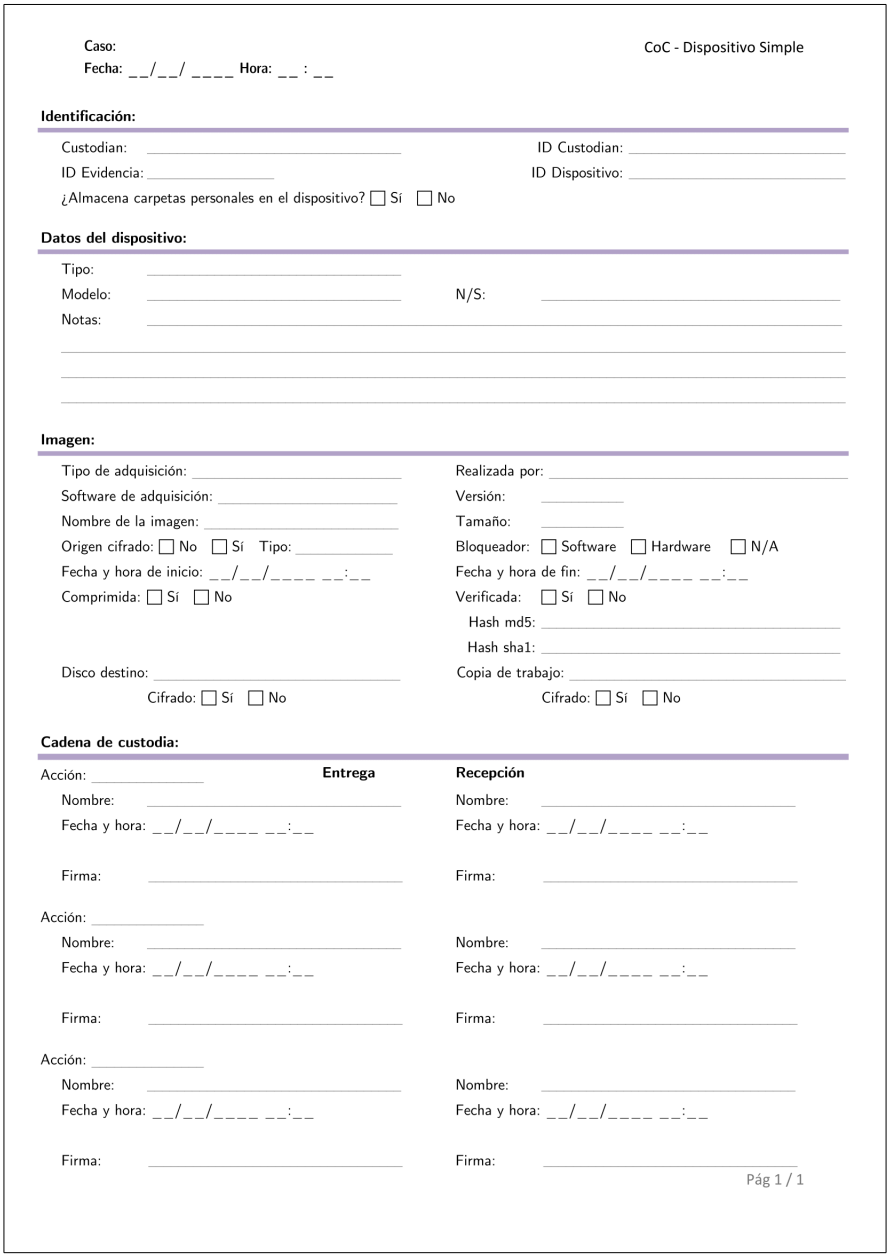
Fecha y hora: __/__/____ __:__ Fecha y hora: __/__/____ __:__

Firma: _____ Firma: _____

Pág 1 / 1

Figura A9: Formulario Cadena de Custodia - Extracción de Información.

A.7. CoC - Dispositivo Simple



Caso:

CoC - Dispositivo Simple

Fecha: __/__/ ____ Hora: __ : __

Identificación:

Custodian: _____ ID Custodian: _____

ID Evidencia: _____ ID Dispositivo: _____

¿Almacena carpetas personales en el dispositivo? ☐ Sí ☐ No

Datos del dispositivo:

Tipo: _____

Modelo: _____ N/S: _____

Notas: _____

Imagen:

Tipo de adquisición: _____ Realizada por: _____

Software de adquisición: _____ Versión: _____

Nombre de la imagen: _____ Tamaño: _____

Origen cifrado: ☐ No ☐ Sí Tipo: _____ Bloqueador: ☐ Software ☐ Hardware ☐ N/A

Fecha y hora de inicio: __/__/____ __:__ Fecha y hora de fin: __/__/____ __:__

Comprimida: ☐ Sí ☐ No Verificada: ☐ Sí ☐ No

Hash md5: _____

Hash sha1: _____

Disco destino: _____ Copia de trabajo: _____

Cifrado: ☐ Sí ☐ No Cifrado: ☐ Sí ☐ No

Cadena de custodia:

Acción: _____ **Entrega** **Recepción**

Nombre: _____ Nombre: _____

Fecha y hora: __/__/____ __:__ Fecha y hora: __/__/____ __:__

Firma: _____ Firma: _____

Acción: _____

Nombre: _____ Nombre: _____

Fecha y hora: __/__/____ __:__ Fecha y hora: __/__/____ __:__

Firma: _____ Firma: _____

Acción: _____

Nombre: _____ Nombre: _____

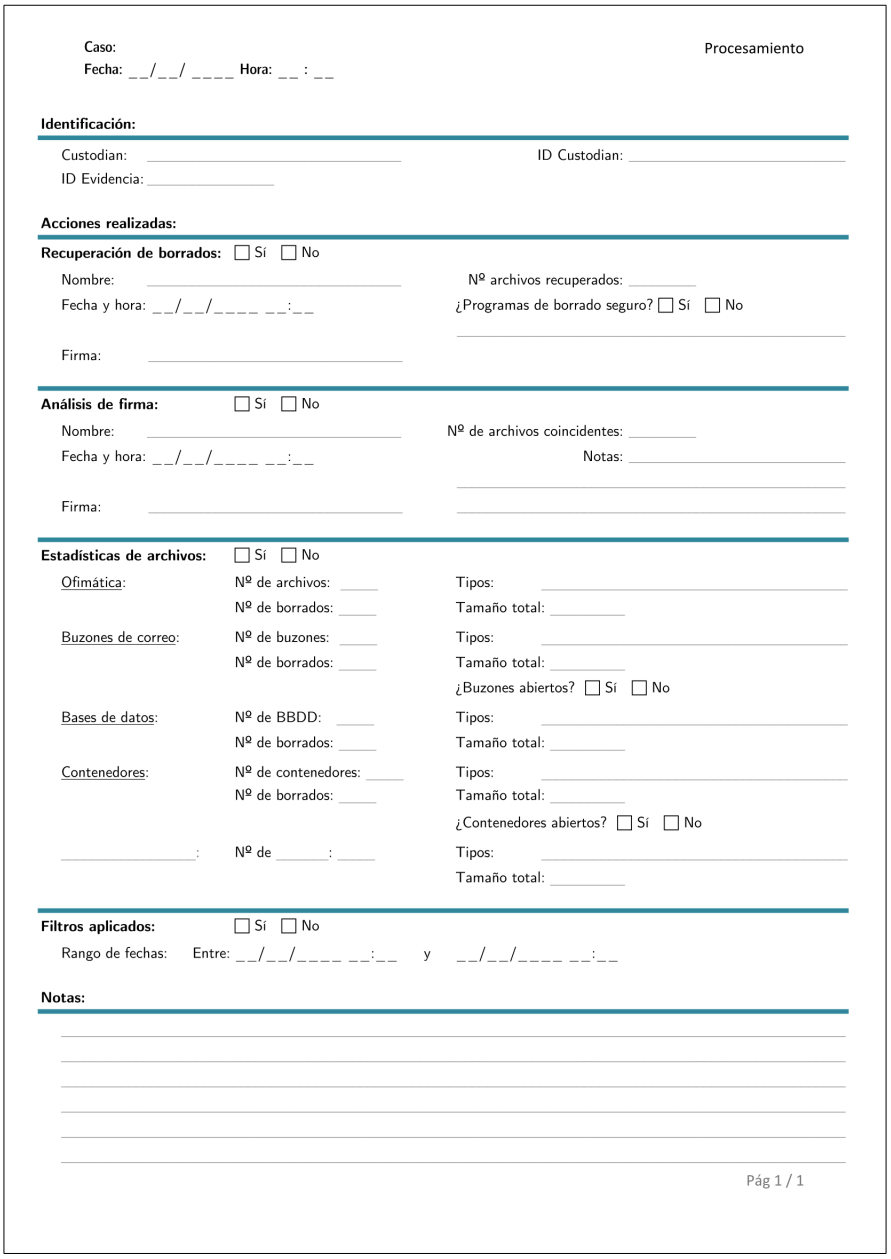
Fecha y hora: __/__/____ __:__ Fecha y hora: __/__/____ __:__

Firma: _____ Firma: _____

Pág 1 / 1

Figura A10: Formulario Cadena de Custodia - Dispositivo Simple.

A.8. Procesamiento



Caso:

Procesamiento

Fecha: __/__/ ____ Hora: __ : __

Identificación:

Custodian: _____ ID Custodian: _____

ID Evidencia: _____

Acciones realizadas:

Recuperación de borrados: ☐ Sí ☐ No

Nombre: _____ Nº archivos recuperados: _____

Fecha y hora: __/__/____ __:__ ¿Programas de borrado seguro? ☐ Sí ☐ No

Firma: _____

Análisis de firma: ☐ Sí ☐ No

Nombre: _____ Nº de archivos coincidentes: _____

Fecha y hora: __/__/____ __:__ Notas: _____

Firma: _____

Estadísticas de archivos: ☐ Sí ☐ No

Ofimática: Nº de archivos: _____ Tipos: _____

Nº de borrados: _____ Tamaño total: _____

Buzones de correo: Nº de buzones: _____ Tipos: _____

Nº de borrados: _____ Tamaño total: _____

¿Buzones abiertos? ☐ Sí ☐ No

Bases de datos: Nº de BBDD: _____ Tipos: _____

Nº de borrados: _____ Tamaño total: _____

Contenedores: Nº de contenedores: _____ Tipos: _____

Nº de borrados: _____ Tamaño total: _____

¿Contenedores abiertos? ☐ Sí ☐ No

_____: Nº de _____: _____ Tipos: _____

Tamaño total: _____

Filtros aplicados: ☐ Sí ☐ No

Rango de fechas: Entre: __/__/____ __:__ y __/__/____ __:__

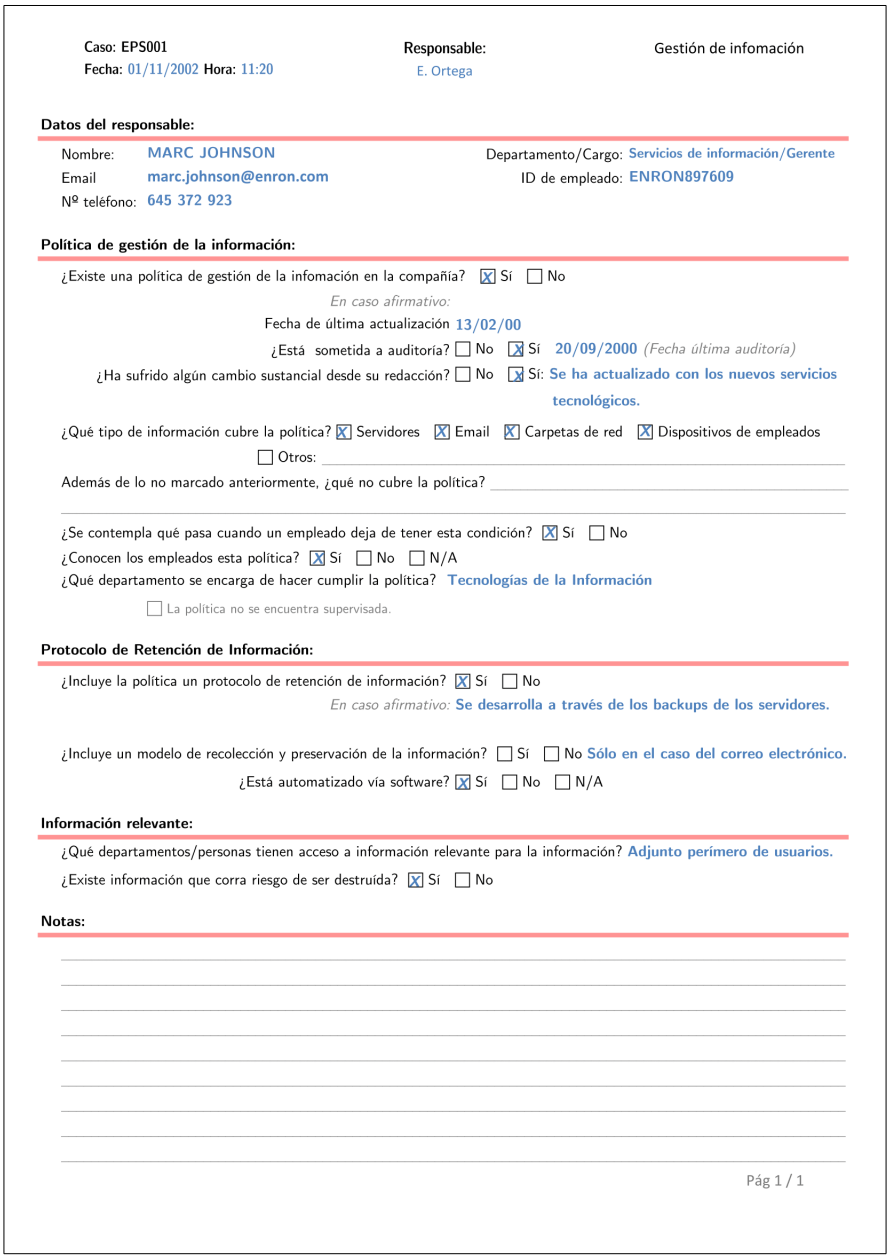
Notas:

Pág 1 / 1

Figura A11: Formulario de Procesamiento.

B. Formularios - Caso práctico

B.1. Gestión de Información - Caso práctico



Caso: EPS001
Fecha: 01/11/2002 Hora: 11:20

Responsable:
E. Ortega

Gestión de infomación

Datos del responsable:

Nombre: MARC JOHNSON
Email marc.johnson@enron.com
Nº teléfono: 645 372 923

Departamento/Cargo: Servicios de información/Gerente
ID de empleado: ENRON897609

Política de gestión de la información:

¿Existe una política de gestión de la infomación en la compañía? [X] Sí [ ] No

En caso afirmativo:

Fecha de última actualización 13/02/00

¿Está sometida a auditoría? [ ] No [X] Sí 20/09/2000 *(Fecha última auditoría)*

¿Ha sufrido algún cambio sustancial desde su redacción? [ ] No [X] Sí: Se ha actualizado con los nuevos servicios tecnológicos.

¿Qué tipo de información cubre la política? [X] Servidores [X] Email [X] Carpetas de red [X] Dispositivos de empleados [ ] Otros: _____

Además de lo no marcado anteriormente, ¿qué no cubre la política? _____

¿Se contempla qué pasa cuando un empleado deja de tener esta condición? [X] Sí [ ] No

¿Conocen los empleados esta política? [X] Sí [ ] No [ ] N/A

¿Qué departamento se encarga de hacer cumplir la política? Tecnologías de la Información

[ ] La política no se encuentra supervisada.

Protocolo de Retención de Información:

¿Incluye la política un protocolo de retención de información? [X] Sí [ ] No

En caso afirmativo: Se desarrolla a través de los backups de los servidores.

¿Incluye un modelo de recolección y preservación de la información? [ ] Sí [ ] No Sólo en el caso del correo electrónico.

¿Está automatizado vía software? [X] Sí [ ] No [ ] N/A

Información relevante:

¿Qué departamentos/personas tienen acceso a información relevante para la información? Adjunto perímero de usuarios.

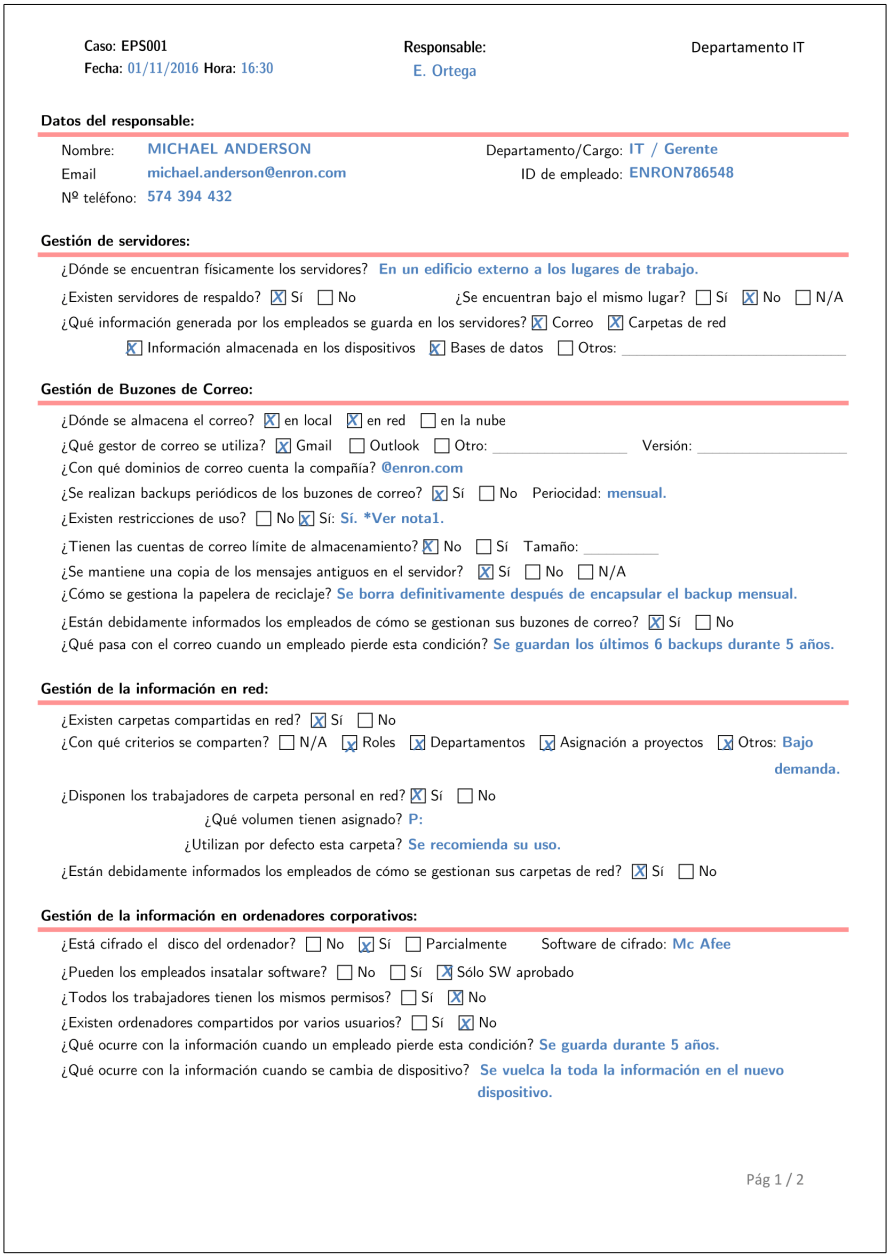
¿Existe información que corra riesgo de ser destruída? [X] Sí [ ] No

Notas:

Pág 1 / 1

Figura B12: Formulario: Gestión de Información.

B.2. Departamento IT - Caso práctico



Caso: EPS001
Fecha: 01/11/2016 Hora: 16:30

Responsable:
E. Ortega

Departamento IT

Datos del responsable:

Nombre: MICHAEL ANDERSON
Email michael.anderson@enron.com
Nº teléfono: 574 394 432

Departamento/Cargo: IT / Gerente
ID de empleado: ENRON786548

Gestión de servidores:

¿Dónde se encuentran fisicamente los servidores? En un edificio externo a los lugares de trabajo.

¿Existen servidores de respaldo? [X] Sí [ ] No ¿Se encuentran bajo el mismo lugar? [ ] Sí [X] No [ ] N/A

¿Qué información generada por los empleados se guarda en los servidores? [X] Correo [X] Carpetas de red [X] Información almacenada en los dispositivos [X] Bases de datos [ ] Otros: _____

Gestión de Buzones de Correo:

¿Dónde se almacena el correo? [X] en local [X] en red [ ] en la nube

¿Qué gestor de correo se utiliza? [X] Gmail [ ] Outlook [ ] Otro: _____ Versión: _____

¿Con qué dominios de correo cuenta la compañía? @enron.com

¿Se realizan backups periódicos de los buzones de correo? [X] Sí [ ] No Periocidad: mensual.

¿Existen restricciones de uso? [ ] No [X] Sí: Sí. *Ver nota1.

¿Tienen las cuentas de correo límite de almacenamiento? [X] No [ ] Sí Tamaño: _____

¿Se mantiene una copia de los mensajes antiguos en el servidor? [X] Sí [ ] No [ ] N/A

¿Cómo se gestiona la papelera de reciclaje? Se borra definitivamente después de encapsular el backup mensual.

¿Están debidamente informados los empleados de cómo se gestionan sus buzones de correo? [X] Sí [ ] No

¿Qué pasa con el correo cuando un empleado pierde esta condición? Se guardan los últimos 6 backups durante 5 años.

Gestión de la información en red:

¿Existen carpetas compartidas en red? [X] Sí [ ] No

¿Con qué criterios se comparten? [ ] N/A [X] Roles [X] Departamentos [X] Asignación a proyectos [X] Otros: Bajo demanda.

¿Disponen los trabajadores de carpeta personal en red? [X] Sí [ ] No

¿Qué volumen tienen asignado? P:

¿Utilizan por defecto esta carpeta? Se recomienda su uso.

¿Están debidamente informados los empleados de cómo se gestionan sus carpetas de red? [X] Sí [ ] No

Gestión de la información en ordenadores corporativos:

¿Está cifrado el disco del ordenador? [ ] No [X] Sí [ ] Parcialmente Software de cifrado: Mc Afee

¿Pueden los empleados insatalar software? [ ] No [ ] Sí [X] Sólo SW aprobado

¿Todos los trabajadores tienen los mismos permisos? [ ] Sí [X] No

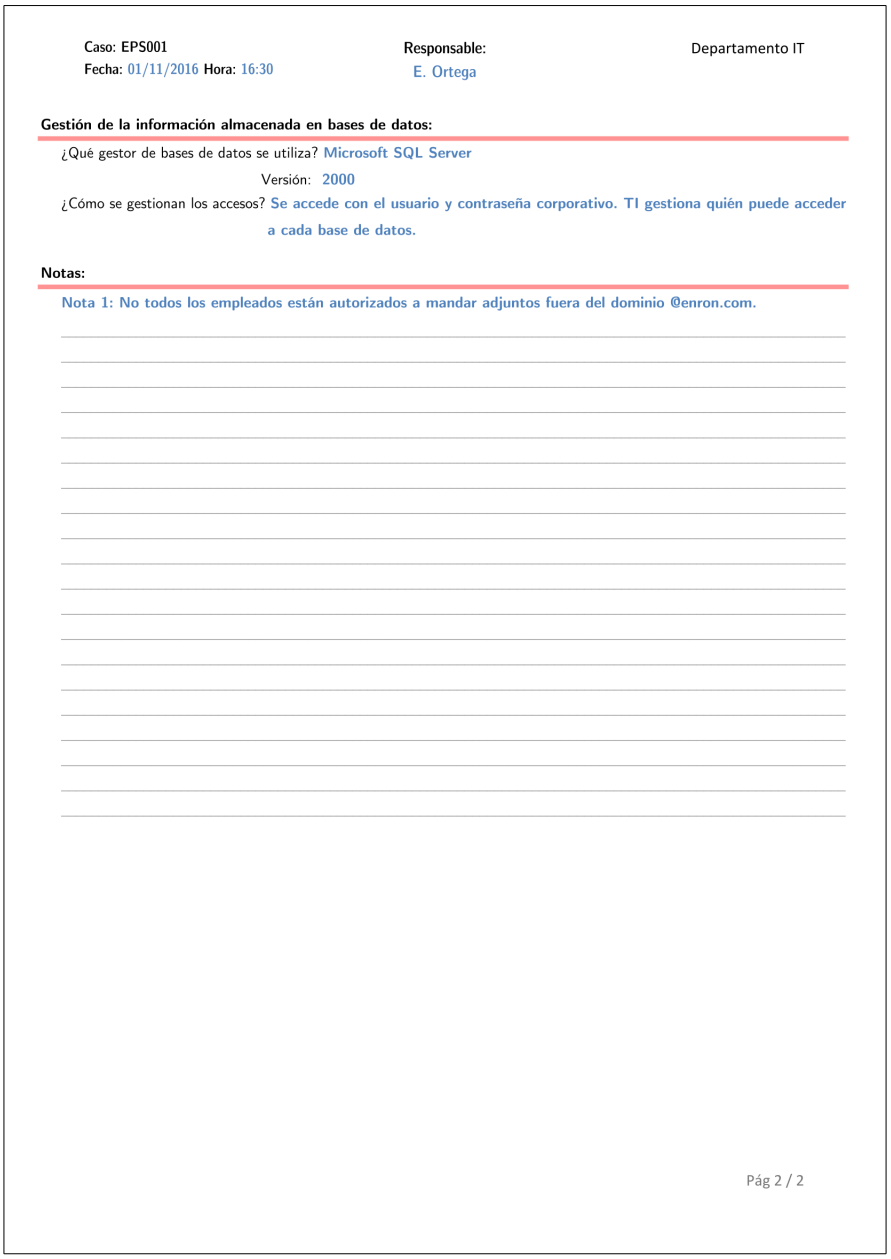
¿Existen ordenadores compartidos por varios usuarios? [ ] Sí [X] No

¿Qué ocurre con la información cuando un empleado pierde esta condición? Se guarda durante 5 años.

¿Qué ocurre con la información cuando se cambia de dispositivo? Se vuelca la toda la información en el nuevo dispositivo.

Pág 1 / 2

Figura B13: Formulario Departamento IT (Anverso).



Caso: EPS001
Fecha: 01/11/2016 Hora: 16:30

Responsable:
E. Ortega

Departamento IT

Gestión de la información almacenada en bases de datos:

¿Qué gestor de bases de datos se utiliza? Microsoft SQL Server

Versión: 2000

¿Cómo se gestionan los accesos? Se accede con el usuario y contraseña corporativo. TI gestiona quién puede acceder a cada base de datos.

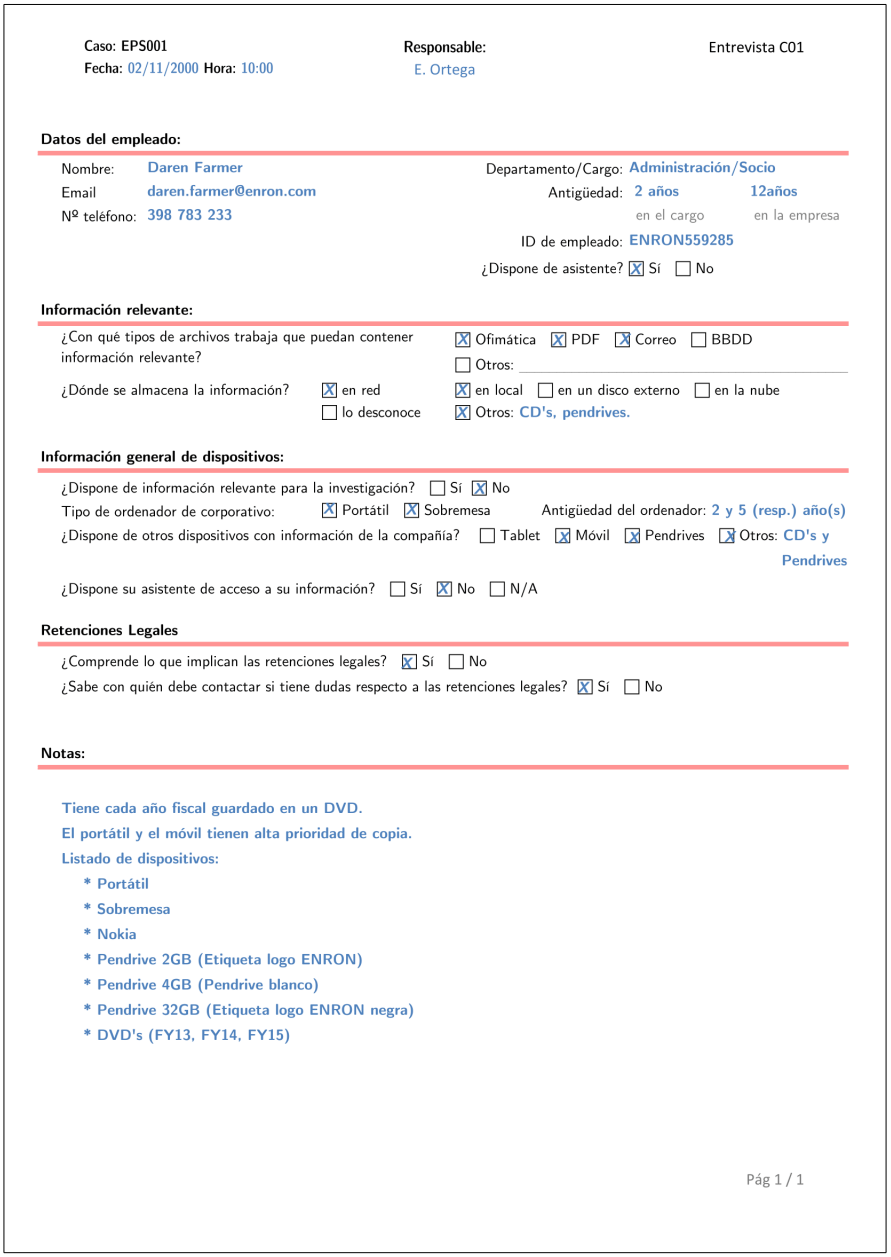
Notas:

Nota 1: No todos los empleados están autorizados a mandar adjuntos fuera del dominio @enron.com.

Pág 2 / 2

Figura B14: Formulario Departamento IT (Reverso).

B.3. Entrevista Custodian - Caso práctico



Caso: EPS001
Fecha: 02/11/2000 Hora: 10:00

Responsable:
E. Ortega

Entrevista C01

Datos del empleado:

Nombre: Daren Farmer
Email daren.farmer@enron.com
Nº teléfono: 398 783 233

Departamento/Cargo: Administración/Socio
Antigüedad: 2 años (en el cargo) 12años (en la empresa)
ID de empleado: ENRON559285
¿Dispone de asistente? [X] Sí [ ] No

Información relevante:

¿Con qué tipos de archivos trabaja que puedan contener información relevante? [X] Ofimática [X] PDF [X] Correo [ ] BBDD [ ] Otros: _____

¿Dónde se almacena la información? [X] en red [ ] lo desconoce [X] en local [ ] en un disco externo [ ] en la nube [X] Otros: CD's, pendrives.

Información general de dispositivos:

¿Dispone de información relevante para la investigación? [ ] Sí [X] No
Tipo de ordenador de corporativo: [X] Portátil [X] Sobremesa Antigüedad del ordenador: 2 y 5 (resp.) año(s)
¿Dispone de otros dispositivos con información de la compañía? [ ] Tablet [X] Móvil [X] Pendrives [X] Otros: CD's y Pendrives

¿Dispone su asistente de acceso a su información? [ ] Sí [X] No [ ] N/A

Retenciones Legales

¿Comprende lo que implican las retenciones legales? [X] Sí [ ] No
¿Sabe con quién debe contactar si tiene dudas respecto a las retenciones legales? [X] Sí [ ] No

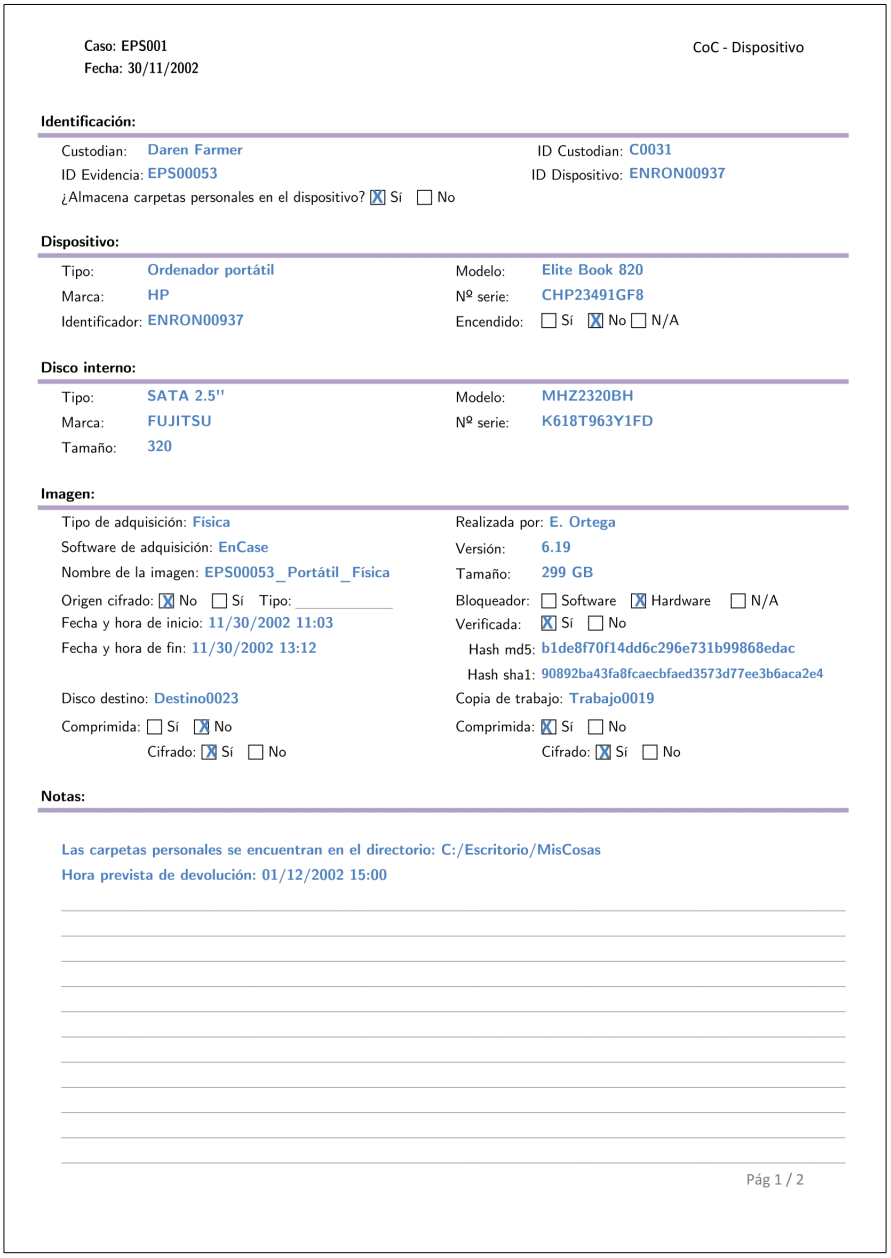
Notas:

Tiene cada año fiscal guardado en un DVD.
El portátil y el móvil tienen alta prioridad de copia.
Listado de dispositivos:
* Portátil
* Sobremesa
* Nokia
* Pendrive 2GB (Etiqueta logo ENRON)
* Pendrive 4GB (Pendrive blanco)
* Pendrive 32GB (Etiqueta logo ENRON negra)
* DVD's (FY13, FY14, FY15)

Pág 1 / 1

Figura B15: Formulario Entrevista al custodian Daren Farmer.

B.4. Información del dispositivo - Caso práctico



Caso: EPS001
Fecha: 30/11/2002

CoC - Dispositivo

Identificación:

Custodian: Daren Farmer
ID Custodian: C0031
ID Evidencia: EPS00053
ID Dispositivo: ENRON00937
¿Almacena carpetas personales en el dispositivo? [X] Sí [ ] No

Dispositivo:

Tipo: Ordenador portátil
Modelo: Elite Book 820
Marca: HP
Nº serie: CHP23491GF8
Identificador: ENRON00937
Encendido: [ ] Sí [X] No [ ] N/A

Disco interno:

Tipo: SATA 2.5"
Modelo: MHZ2320BH
Marca: FUJITSU
Nº serie: K618T963Y1FD
Tamaño: 320

Imagen:

Tipo de adquisición: Física
Realizada por: E. Ortega
Software de adquisición: EnCase
Versión: 6.19
Nombre de la imagen: EPS00053_Portátil_Física
Tamaño: 299 GB
Origen cifrado: [X] No [ ] Sí Tipo: _____
Bloqueador: [ ] Software [X] Hardware [ ] N/A
Fecha y hora de inicio: 11/30/2002 11:03
Verificada: [X] Sí [ ] No
Fecha y hora de fin: 11/30/2002 13:12
Hash md5: b1de8f70f14dd6c296e731b99868edac
Hash sha1: 90892ba43fa8fcaecbfaed3573d77ee3b6aca2e4
Disco destino: Destino0023
Copia de trabajo: Trabajo0019
Comprimida: [ ] Sí [X] No
Comprimida: [X] Sí [ ] No
Cifrado: [X] Sí [ ] No
Cifrado: [X] Sí [ ] No

Notas:

Las carpetas personales se encuentran en el directorio: C:/Escritorio/MisCosas
Hora prevista de devolución: 01/12/2002 15:00

Pág 1 / 2

Figura B16: Formulario Cadena de Custodia - Dispositivo (Anverso).

Caso: EPS001
Fecha: 30/11/2002

CoC - Dispositivo

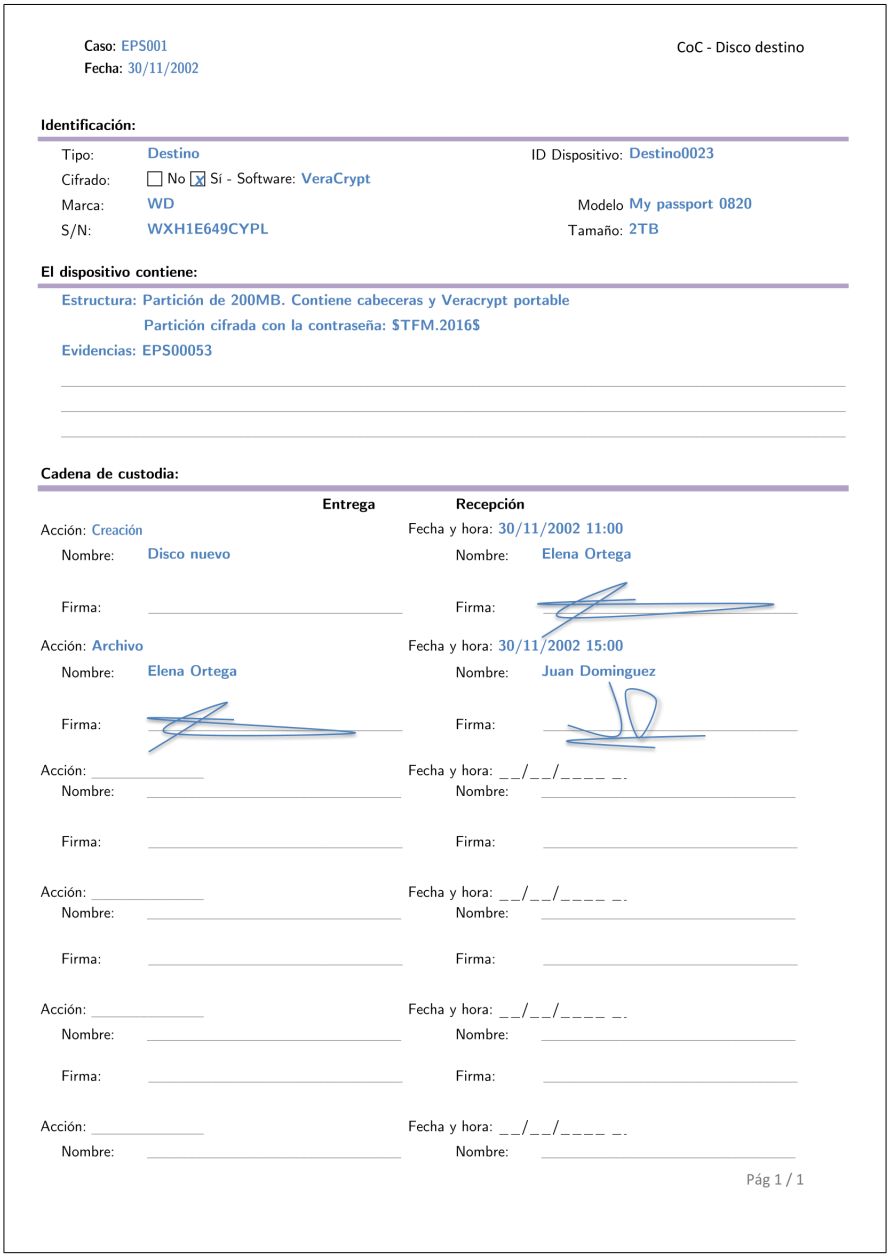
Cadena de custodia:

	Entrega	**Recepción**	
c		Fecha y hora: 30/11/2002 10:15	
Nombre:	Daren Farmer	Nombre:	Elena Ortega
Firma:		Firma:	
Acción: Devolución del dispositivo.		Fecha y hora: 01/12/2002 15:15	
Nombre:	Elena Ortega	Nombre:	Daren Farmer
Firma:		Firma:	
Acción: ______		Fecha y hora: __/__/____ __	
Nombre:		Nombre:	
Firma:		Firma:	
Acción: ______		Fecha y hora: __/__/____ __	
Nombre:		Nombre:	
Firma:		Firma:	
Acción: ______		Fecha y hora: __/__/____ __	
Nombre:		Nombre:	
Firma:		Firma:	
Acción: ______		Fecha y hora: __/__/____ __	
Nombre:		Nombre:	
Firma:		Firma:	
Acción: ______		Fecha y hora: __/__/____ __	
Nombre:		Nombre:	
Firma:		Firma:	
Acción: ______		Fecha y hora: __/__/____ __	
Nombre:		Nombre:	
Firma:		Firma:	
Acción: ______		Fecha y hora: __/__/____ __	
Nombre:		Nombre:	
Firma:		Firma:	
Acción: ______		Fecha y hora: __/__/____ __	
Nombre:		Nombre:	
Firma:		Firma:	

Pág 2 / 2

Figura B17: Formulario Cadena de Custodia - Dispositivo (Reverso).

B.5. Disco Destino - Caso práctico



Caso: EPS001
Fecha: 30/11/2002

CoC - Disco destino

Identificación:

Tipo: Destino
ID Dispositivo: Destino0023
Cifrado: ☐ No ☒ Sí - Software: VeraCrypt
Marca: WD
Modelo My passport 0820
S/N: WXH1E649CYPL
Tamaño: 2TB

El dispositivo contiene:

Estructura: Partición de 200MB. Contiene cabeceras y Veracrypt portable
Partición cifrada con la contraseña: $TFM.2016$
Evidencias: EPS00053

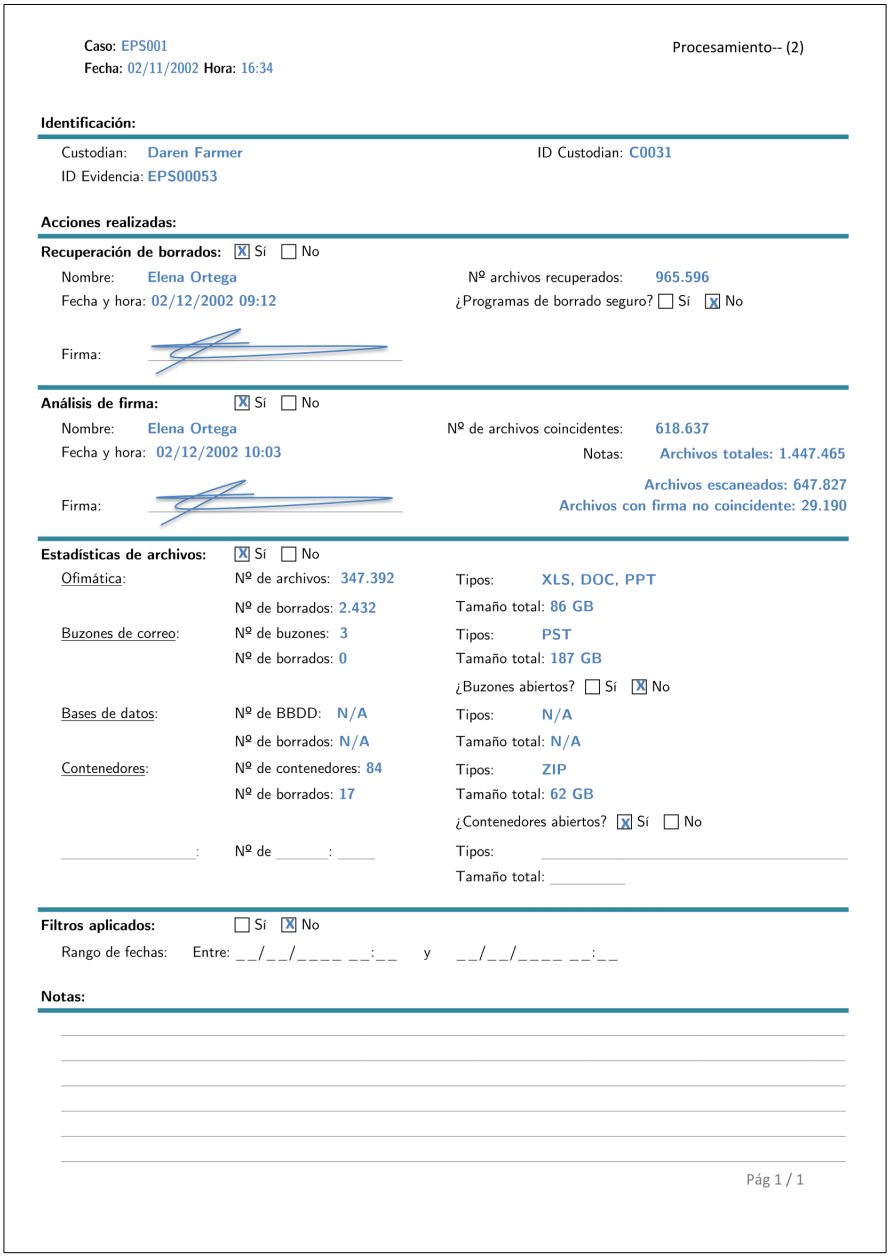
Cadena de custodia:

	Entrega	Recepción
Acción: Creación		Fecha y hora: 30/11/2002 11:00
Nombre:	Disco nuevo	Elena Ortega
Firma:		
Acción: Archivo		Fecha y hora: 30/11/2002 15:00
Nombre:	Elena Ortega	Juan Dominguez
Firma:		
Acción:		Fecha y hora: __/__/____ _.
Nombre:		
Firma:		
Acción:		Fecha y hora: __/__/____ _.
Nombre:		
Firma:		
Acción:		Fecha y hora: __/__/____ _.
Nombre:		
Firma:		
Acción:		Fecha y hora: __/__/____ _.
Nombre:		

Pág 1 / 1

Figura B18: Formulario Disco Destino.

B.6. Procesamiento - Caso práctico



Caso: EPS001
Fecha: 02/11/2002 Hora: 16:34

Procesamiento-- (2)

Identificación:

Custodian: Daren Farmer
ID Custodian: C0031
ID Evidencia: EPS00053

Acciones realizadas:

Recuperación de borrados: [X] Sí [ ] No
Nombre: Elena Ortega
Nº archivos recuperados: 965.596
Fecha y hora: 02/12/2002 09:12
¿Programas de borrado seguro? [ ] Sí [X] No
Firma:

Análisis de firma: [X] Sí [ ] No
Nombre: Elena Ortega
Nº de archivos coincidentes: 618.637
Fecha y hora: 02/12/2002 10:03
Notas: Archivos totales: 1.447.465
Archivos escaneados: 647.827
Archivos con firma no coincidente: 29.190
Firma:

Estadísticas de archivos: [X] Sí [ ] No

Ofimática:
Nº de archivos: 347.392
Tipos: XLS, DOC, PPT
Nº de borrados: 2.432
Tamaño total: 86 GB

Buzones de correo:
Nº de buzones: 3
Tipos: PST
Nº de borrados: 0
Tamaño total: 187 GB
¿Buzones abiertos? [ ] Sí [X] No

Bases de datos:
Nº de BBDD: N/A
Tipos: N/A
Nº de borrados: N/A
Tamaño total: N/A

Contenedores:
Nº de contenedores: 84
Tipos: ZIP
Nº de borrados: 17
Tamaño total: 62 GB
¿Contenedores abiertos? [X] Sí [ ] No

_____: Nº de _____: ____
Tipos: _____
Tamaño total: _____

Filtros aplicados: [ ] Sí [X] No
Rango de fechas: Entre: __/__/____ __:__ y __/__/____ __:__

Notas:

Pág 1 / 1

Figura B19: Formulario Procesamiento.

C. Report EnCase

Name	EV_00001_Portatil_Fisica
In Report	No
Description	Physical Disk, 625.142.447 Sectors 298,1GB
Is Deleted	No
Logical Size	0
Initialized Size	0
Physical Size	512
Starting Extent	0S0
File Extents	1
Permissions	No
References	0
Physical Location	0
Physical Sector	0
Evidence File	EV_00001_Portatil_Fisica
File Identifier	0
Code Page	0
Hash Properties	No
Full Path	TFM\EV_00001_Portatil_Fisica
Is Duplicate	No
Is Internal	No
Is Overwritten	No

Device

Name	EV_00001_Portatil_Fisica
Actual Date	05/08/2016 16:01:31
Target Date	05/08/2016 16:01:31
File Path	D:\TFM-ELENA\EV_00001\EV_00001_Portatil_Fisica.E01
Case Number	TFM
Evidence Number	EV_00001
Examiner Name	EOH
Notes	Custodian: Bárcena Beltrán, José Francisco; Dispositivo: Tipo: Portátil; Modelo...
Label	Seagate
Model	Expansion
Serial Number	#" ÂExpansion
Drive Type	Fixed
File Integrity	Completely Verified, 0 Errors
Acquisition MD5	b1de8f70f14dd6c296e731b99868edac
Verification MD5	b1de8f70f14dd6c296e731b99868edac
Acquisition SHA1	90892ba43fa8fcaecbfaed3573d77ee3b6aca2e4
Verification SHA1	90892ba43fa8fcaecbfaed3573d77ee3b6aca2e4
GUID	bdee84539118bd47afd78a8394cdd017
EnCase Version	6.18
System Version	Windows 7
Neutrino	No
Is Physical	Yes
Raid RHS	No
Raid Stripe Size	0
Error Granularity	64
Process ID	0
Index File	C:\Program Files\EnCase6\Index\EV_00001_Portatil_Fisica-bdee84539118bd47afd78a8394cdd017.Index
Acquisition Info	No
Sources	No
Subjects	No
Read Errors	0

Figura C20: Report EnCase (Anverso).

Missing Sectors	0
Disk Elements	No
CRC Errors	0
Compression	None
Total Size	320.072.932.864 Bytes (298,1GB)
Total Sectors	625.142.447
Disk Signature	E1B750AE
Partitions	Valid

Hash Properties

Name	Value
Hash Set	
Hash Category	

Partitions

Name	Id	Type	Start Sector	Total Sectors	Size
	07	NTFS	2.048	204.800	100MB
	07	NTFS	208.845	564.737.087	269,3GB
	83	Linux Native	564.946.881	57.620.480	27,5GB
	82	Linux Swap	622.569.409	2.572.288	1,2GB

Figura C21: Report EnCase (Reverso).

D. Report Keywords

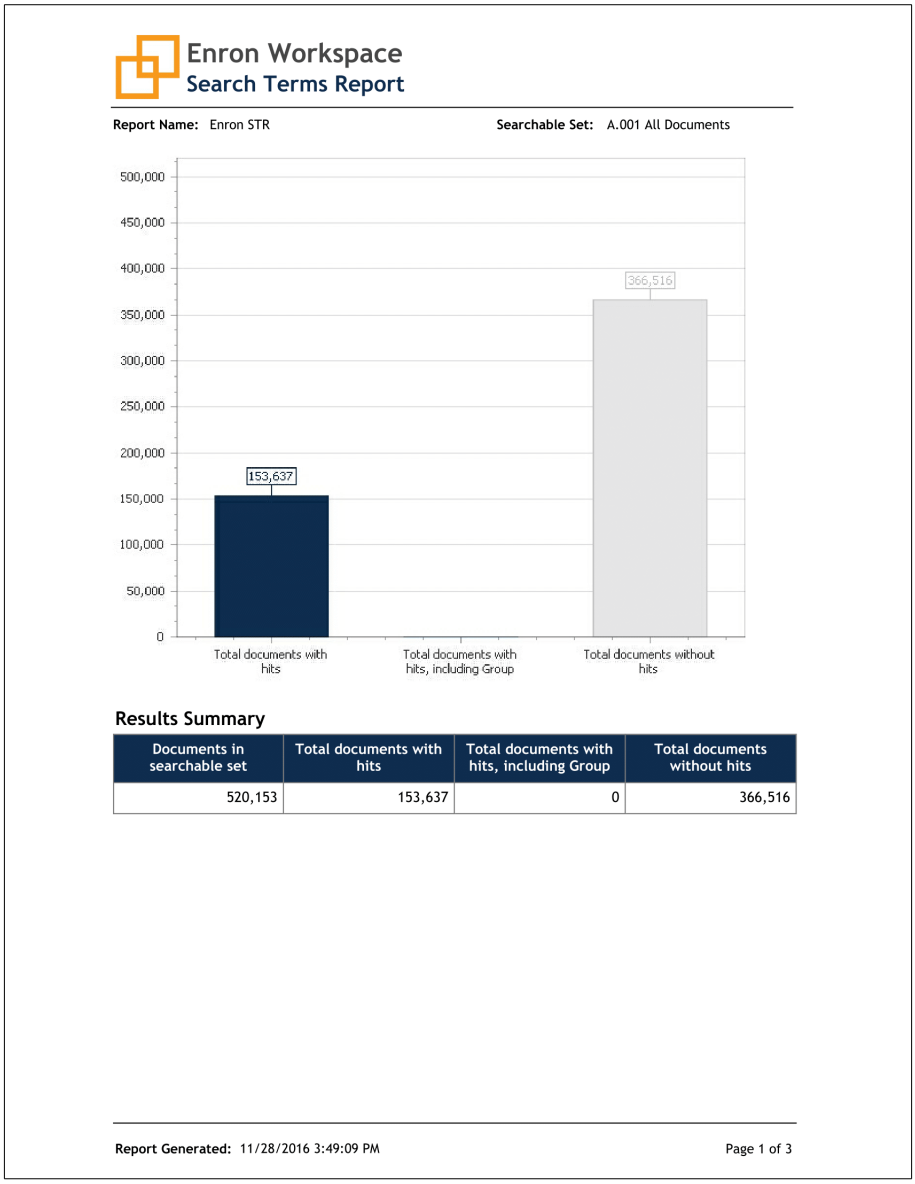


Figura D22: Relativity: Report Keywords (Página 1/3).

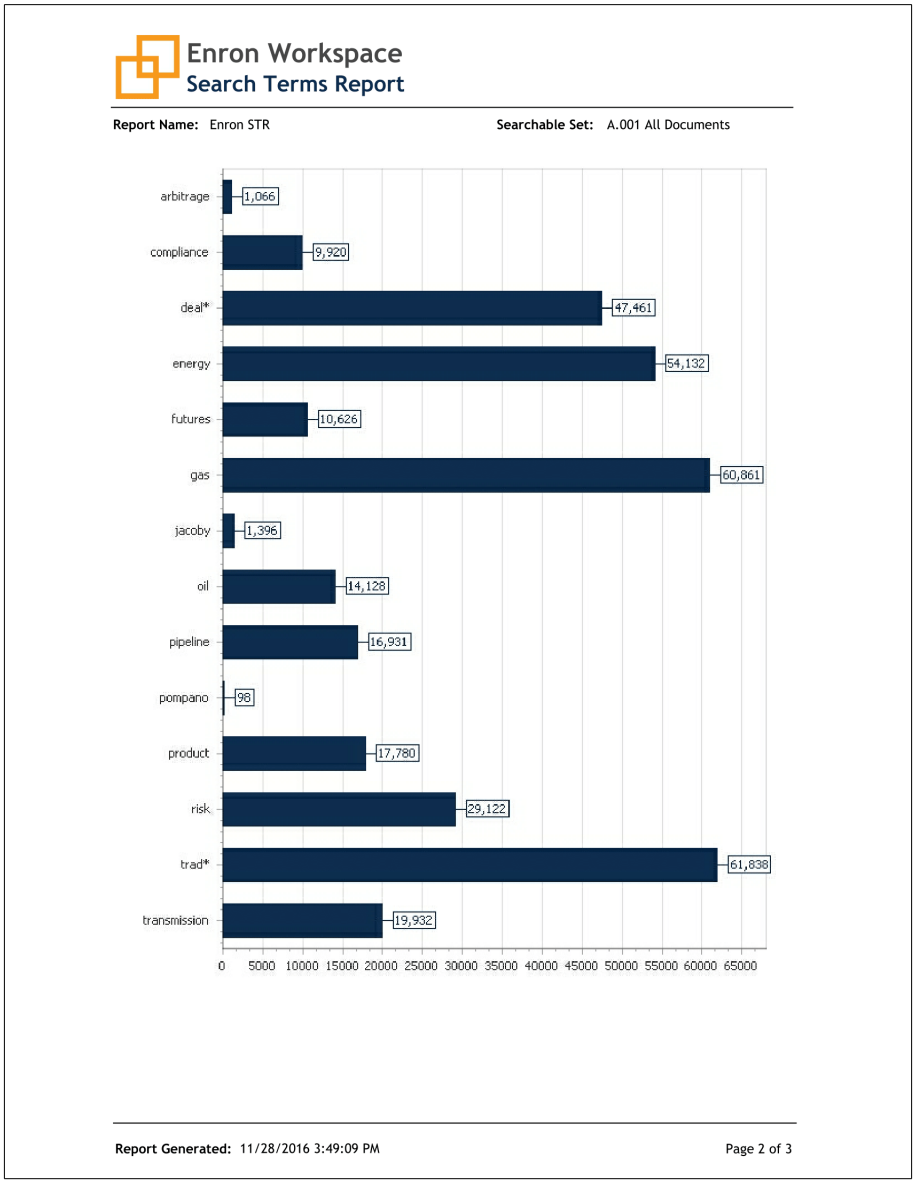
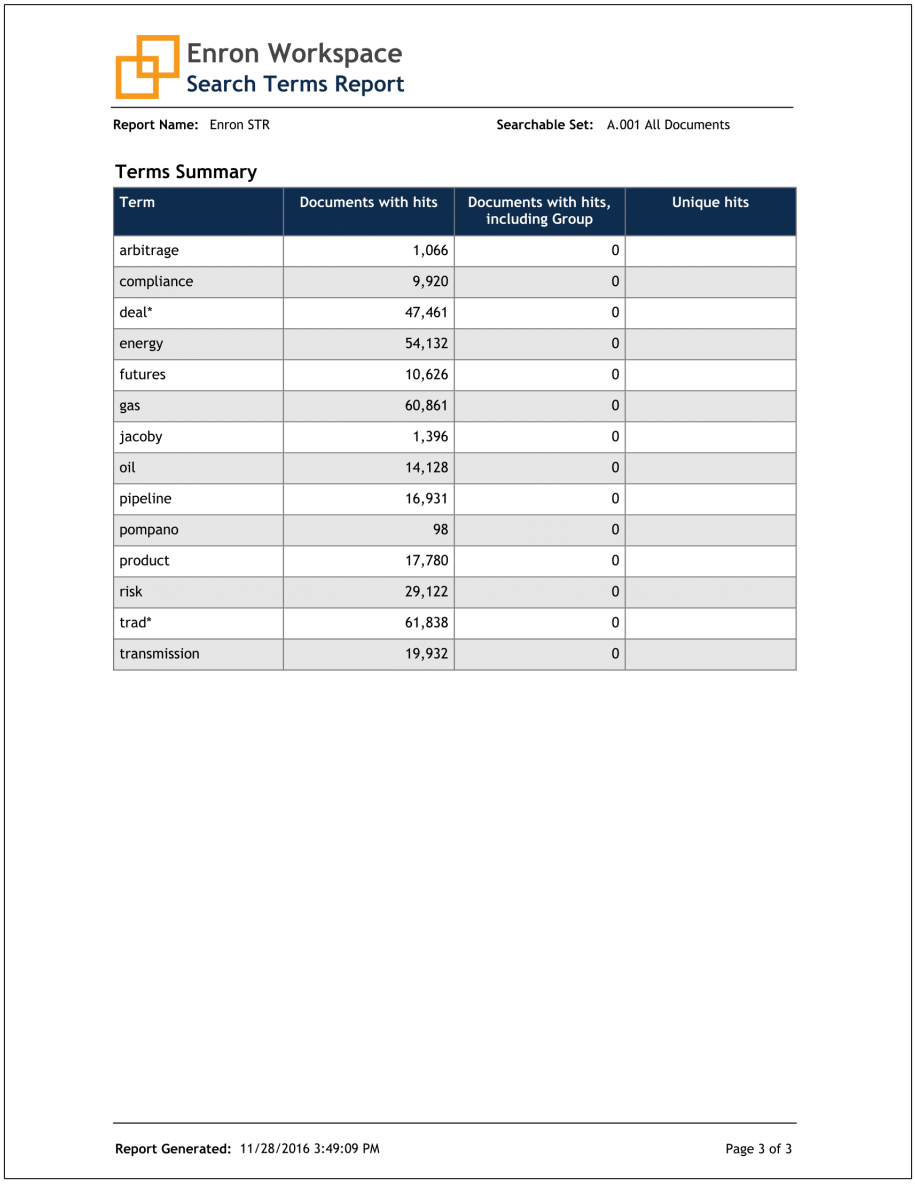


Figura D23: Relativity: Report Keywords (Página 2/3).



Enron Workspace
Search Terms Report

Report Name: Enron STR **Searchable Set:** A.001 All Documents

Terms Summary

Term	Documents with hits	Documents with hits, including Group	Unique hits
arbitrage	1,066	0	
compliance	9,920	0	
deal*	47,461	0	
energy	54,132	0	
futures	10,626	0	
gas	60,861	0	
jacoby	1,396	0	
oil	14,128	0	
pipeline	16,931	0	
pompano	98	0	
product	17,780	0	
risk	29,122	0	
trad*	61,838	0	
transmission	19,932	0	

Report Generated: 11/28/2016 3:49:09 PM Page 3 of 3

Figura D24: Relativity: Report Keywords (Página 3/3).