

Nicolle M. Llord Ciriello

**DESARROLLO DE UNA SOLUCIÓN CON AZURE PARA LA GESTIÓN Y
PROTECCIÓN DE DATOS**

TRABAJO DE FIN DE GRADO

dirigido por Usama Benabdelkrim

Grado de Ingeniería Informática



UNIVERSITAT ROVIRA I VIRGILI

Tarragona 2025

Resum.

La creixent necessitat de protegir informació sensible a les organitzacions ha motivat el desenvolupament d'aquest treball de fi de grau, que té com objectiu implementar una solució integral al núvol d'Azure per gestionar i protegir les dades d'un despatx fictici d'advocats especialitzat en patents i marques. Aquesta solució pretén garantir la seguretat, la governança, la confidencialitat i el compliment normatiu en el tractament de la informació.

El treball s'estructura en tres parts amb resultats diferenciats:

1. **Governança de dades:** s'ha implementat una solució híbrida d'identitat mitjançant Microsoft Entra ID Connect, que permet l'ús d'inici de sessió únic i una administració híbrida d'usuaris entre entorns locals i al núvol.
2. **Protecció de dades:** s'ha dut a terme la classificació i xifrat de la informació mitjançant etiquetes definides amb Microsoft Information Protection, així com la configuració de polítiques de prevenció de pèrdua de dades Data Loss Prevention i la gestió de riscos interns a través de polítiques d'Insider Risk Management. També s'ha configurat la gestió i protecció del cicle de vida de les dades mitjançant Data Lifecycle Management.
3. **Desenvolupament d'aplicació web:** per integrar tota la solució de protecció i gestió de dades, s'ha desenvolupat una aplicació web que permet als clients concertar cites amb el despatx d'advocats. Aquesta aplicació compta amb un backend desenvolupat amb Spring Boot i un frontend amb React i TypeScript. El flux d'ús de l'aplicació permet als clients fer reserves de cites i generar acords de confidencialitat i factures segons el servei sol·licitat. Les dades generades són posteriorment classificades i protegides mitjançant tota la infraestructura de protecció i governança configurada prèviament.

Resumen.

La creciente necesidad de proteger información sensible en organizaciones ha motivado el desarrollo de este trabajo de fin de grado, cuyo objetivo es implementar una solución integral en la nube de Azure para gestionar y proteger los datos de un despacho ficticio de abogados especializado en patentes y marcas. Esta solución busca garantizar la seguridad, gobernanza, confidencialidad y cumplimiento normativo en el tratamiento de la información.

El trabajo consta de tres partes y resultados diferenciados:

1. **Gobernanza de datos:** se ha implementado una solución híbrida de identidad mediante Microsoft Entra ID Connect, que permite el uso de inicio de sesión único (Single Sign-On) y una administración de usuarios híbrida entre entornos locales y en la nube.
2. **Protección de datos:** se ha llevado a cabo la clasificación y el cifrado de la información mediante etiquetas definidas con Microsoft Information Protection, así como la configuración de políticas de prevención de pérdida de datos de Data Loss Prevention y gestión de riesgos internos a través de políticas de Insider Risk Management. También se ha configurado la gestión y protección del ciclo de vida de los datos con Data Lifecycle Management.

3. **Desarrollo de aplicación web:** para englobar toda la solución de protección y gestión de datos se ha creado una web para agendar citas por parte de los clientes en el despacho de abogados. Esta web cuenta con un backend desarrollado en Springboot y un frontend desarrollado en React con TypeScript. El flujo de uso de la aplicación web permite realizar reservas de citas a los clientes y generar acuerdos de confidencialidad y facturas en base al servicio solicitado. Estos datos generados posteriormente son clasificados y protegidos con toda la infraestructura de protección y gestión de datos previamente configurada.

Abstract.

The growing need to protect sensitive information within organizations has motivated the development of this final degree project, which aims to implement a comprehensive cloud-based solution using Microsoft Azure to manage and safeguard the data of a fictional law firm specializing in patents and trademarks. This solution seeks to ensure security, governance, confidentiality, and regulatory compliance in the handling of information.

The project is structured into three main components, each with distinct outcomes:

1. **Data Governance:** A hybrid identity solution has been implemented using Microsoft Entra ID Connect, enabling Single Sign-On and hybrid user management across both on-premises and cloud environments.
2. **Data Protection:** Information has been classified and encrypted using labels defined with Microsoft Information Protection. Additionally, policies for Data Loss Prevention (DLP) and Insider Risk Management have been configured to prevent data leaks and address internal threats. Data Lifecycle Management has also been set up to manage and protect the entire data lifecycle.
3. **Web Application Development:** To unify the entire data protection and management framework, a web application has been developed that allows clients to schedule appointments at the law firm. The application features a backend built with Spring Boot and a frontend developed in React with TypeScript. Through the app's workflow, clients can book appointments and automatically generate confidentiality agreements and invoices based on the selected legal service. All generated data is subsequently classified and protected using the previously configured protection and governance infrastructure.

1	INTRODUCCIÓN	7
2	PALABRAS CLAVE DEL PROYECTO	8
3	DESCRIPCIÓN GENERAL DEL PROYECTO	9
3.1	NECESIDADES	9
3.1.1	<i>Organización de documentos locales</i>	9
3.1.2	<i>Cumplimiento de periodos de retención de los documentos</i>	10
3.1.3	<i>Protección de la información confidencial</i>	10
3.1.4	<i>Migración de usuarios e inicio de sesión único</i>	10
3.1.5	<i>Monitorización de riesgos internos</i>	11
3.1.6	<i>Gestión de citas de forma automatizada</i>	11
3.2	PREVISIONES DE USO	11
3.2.1	<i>Uso de scripts para los documentos locales</i>	11
3.2.2	<i>Uso de scripts para el cumplimiento del periodo de retención</i>	12
3.2.3	<i>Uso de etiquetas de sensibilidad</i>	12
3.2.4	<i>Inicio de sesión único para los empleados</i>	12
3.2.5	<i>Gestión centralizada para los administradores del sistema</i>	12
3.2.6	<i>Reserva de citas y generación de documentos</i>	12
3.2.7	<i>Revisión de alertas de riesgos internos</i>	12
4	OBJETIVOS	13
4.1	OBJETIVO GENERAL	13
4.2	OBJETIVOS ESPECÍFICOS	13
4.3	JUSTIFICACIÓN DE LA ELECCIÓN DEL TEMA	14
5	PLANIFICACIÓN	15
6	REQUISITOS	16
6.1	REQUISITOS FUNCIONALES	16
6.1.1	<i>Fase identidad híbrida</i>	16
6.1.2	<i>Fase protección de datos</i>	19
6.1.3	<i>Fase aplicación web</i>	30
7	REQUISITOS NO FUNCIONALES	36
8	TECNOLOGÍAS EMPLEADAS EN PROTECCIÓN DE DATOS: MICROSOFT PURVIEW	37
8.1	ETIQUETAS DE SENSIBILIDAD (SENSITIVITY LABELS)	37
8.2	TIPOS DE INFORMACIÓN SENSIBLE (SENSITIVE INFO TYPES)	37
8.3	PREVENCIÓN DE PÉRDIDA DE DATOS (DATA LOSS PREVENTION - DLP)	37
8.4	DATA LIFECYCLE MANAGEMENT (DLM)	38
8.5	ADAPTIVE SCOPES	38
8.6	REFINABLE STRINGS (PROPIEDADES GESTIONABLES EN SHAREPOINT)	38
8.6.1	<i>Crawled Properties</i>	39
8.7	INSIDER RISK MANAGEMENT (IRM)	39
8.7.1	<i>Eventos desencadenantes (Triggering Events)</i>	39
8.7.2	<i>Indicadores de riesgo</i>	40
8.7.3	<i>Clasificación de alertas y detección inteligente</i>	40
9	DISEÑO	41
9.1	DISEÑO FASE USUARIOS HÍBRIDOS	41
9.2	DISEÑO DE USUARIOS Y GRUPOS DE M365	41
9.2.1	<i>Diseño departamento de usuarios y licencias</i>	41
9.2.2	<i>Diseño de grupos</i>	41

9.3	DISEÑO FASE PROTECCIÓN DE DATOS	42
9.3.1	<i>Microsoft Sensitive Info Types</i>	42
9.3.2	<i>Microsoft Information Protection</i>	43
9.3.3	<i>Data Loss Prevention</i>	44
9.3.4	<i>Insider Risk Management</i>	45
9.3.5	<i>Data Lifecycle Management con Adaptive Scopes</i>	45
9.4	DISEÑO APLICACIÓN WEB.....	47
9.4.1	<i>Backend</i>	47
9.4.2	<i>Frontend</i>	47
10	IMPLEMENTACIÓN.....	53
10.1	FASE DE USUARIOS HÍBRIDOS	53
10.1.1	<i>Creación de una red local</i>	53
10.1.2	<i>Implementación de usuarios híbridos</i>	54
10.2	FASE DE PROTECCIÓN DE DATOS	57
10.2.1	<i>Sensitive Info Types</i>	57
10.2.2	<i>Microsoft Information Protection</i>	62
10.2.3	<i>Data loss Prevention</i>	68
10.2.4	<i>Insider Risk Management</i>	70
10.2.5	<i>Datalifecycle Management</i>	85
10.3	FASE DE APLICACIÓN WEB	94
10.3.1	<i>Backend</i>	94
10.3.2	<i>Frontend</i>	98
11	EVALUACIÓN.....	101
11.1	FASE USUARIOS HÍBRIDOS	101
11.1.1	<i>Juego de pruebas extra</i>	104
11.2	FASE PROTECCIÓN DE DATOS	104
11.2.1	<i>Microsoft Information Protection</i>	105
11.2.2	<i>Insider Risk Management</i>	110
11.2.3	<i>Datalifecycle Management</i>	125
11.3	FASE APLICACIÓN WEB.....	128
11.3.1	<i>Pruebas Backend</i>	128
11.3.2	<i>Pruebas Frontend</i>	129
12	EVALUACIÓN DE COSTOS	130
12.1	COSTO DEL PERSONAL.....	130
12.2	COSTO DEL SOFTWARE Y HARDWARE	130
12.3	COSTO TOTAL.....	130
13	VALORACIÓN PERSONAL.....	131
14	ANEXOS	132
15	BIBLIOGRAFÍA	133

Índice de tablas

TABLA 1. PRUEBAS USUARIOS HÍBRIDOS.....	104
TABLA 2. PRUEBAS ADICIONALES MICROSOFT INFORMATION PROTECTION	110
TABLA 3. PRUEBAS SCRIPTS DLM	127
TABLA 4. PRUEBAS BACKEND	128
TABLA 5. PRUEBAS FRONTEND.....	129

Índice de figuras

FIGURA 1. DIAGRAMA DE GANTT, TAREAS 1-34	15
FIGURA 2. DIAGRAMA DE GANTT, TAREAS 34-66	15
FIGURA 3. DIAGRAMA DE GANTT, TAREAS 52-82	15
FIGURA 4. DIAGRAMA CASOS DE USO FASE IDENTIDAD HÍBRIDA.....	16
FIGURA 5. DIAGRAMA DE CASOS DE USO MICROSOFT INFORMATION PROTECTION	22
FIGURA 6. DIAGRAMA DE CASOS DE USO INSIDER RISK MANAGEMENT.....	24
FIGURA 7. DIAGRAMA DE CASOS DE USO DATALIFECYCLE MANAGEMENT.....	27
FIGURA 8. DIAGRAMA DE CASOS DE USO DE LA APLICACIÓN WEB.....	31
FIGURA 9. VISTA DE INICIO.....	48
FIGURA 10. VISTA DE INICIO CON EL CURSOR SOBRE EL TRÁMITE DESEADO	49
FIGURA 11. VISTA DE SELECCIÓN DE FRANJA HORARIA	49
FIGURA 12. VISTA DE FORMULARIO.....	50
FIGURA 13. VISTA DE FORMULARIO SIN RELLENAR	50
FIGURA 14. SELECTOR DE PAÍS.....	51
FIGURA 15. VISTA DE CONFIRMACIÓN DE CITA	51
FIGURA 16. VISTA DE CITA REALIZADA CON ÉXITO	52
FIGURA 17. ASPECTO SERVIDOR LOCAL	54
FIGURA 18. HABILITACIÓN TLS1.2	54
FIGURA 19. BIENVENIDA MICROSOFT ENTRA CONNECT.....	55
FIGURA 20. SELECCIÓN DE CONFIGURACIÓN ENTRA CONNECT	55
FIGURA 21. CONECTAR AD DS.....	56
FIGURA 22. CONFIGURACIÓN COMPLETADA.....	56
FIGURA 23. PATRÓN REGISTRY	57
FIGURA 24. RESUMEN PATRÓN REGISTRY.....	58
FIGURA 25. PATRÓN HEADER.....	59
FIGURA 26. PATRÓN FIELDS.....	59
FIGURA 27. RESUMEN PATRÓN HEADER Y FIELDS.....	60
FIGURA 28. REGEX PRE_DIVISA	60
FIGURA 29. REGEX FINAL_DIVISA.....	61
FIGURA 30. LISTA DE PALABRAS CLAVE INVOICE_LIST.....	61
FIGURA 31. RESUMEN DE INFORMACIÓN SENSIBLE “INVOICE”	62
FIGURA 32. RESUMEN DE LA ETIQUETA “CONFIDENTIAL NON ENCRYPTED”	63
FIGURA 33. PARTE 1 CONTROL DE ACCESO CIFRADO.....	63
FIGURA 34. PARTE 2 CONTROL DE ACCESO CIFRADO.....	64
FIGURA 35. DEFINICIÓN DE INFORMACIÓN SENSIBLE ETIQUETA “CONFIDENTIAL ENCRYPTED”	64
FIGURA 36. DEFINICIÓN DE RECOMENDACIÓN ETIQUETA “CONFIDENTIAL ENCRYPTED”	65
FIGURA 37. DETALLE DEL CONTROL DE ACCESO.	65
FIGURA 38. AUTO-LABELIG ETIQUETA INDUSTRIAL PROPERTY USE.....	66
FIGURA 39. DEFINICIÓN INFORMACIÓN SENSIBLE ETIQUETA INDUSTRIAL PROPERTY USE	66
FIGURA 40. CONFIGURACIÓN ETIQUETA INTERNAL USE.....	67
FIGURA 41. CONFIGURACIÓN DE POLÍTICAS DE PUBLICACIÓN DE LAS ETIQUETAS.....	67
FIGURA 42. CONFIGURACIÓN RESUMEN DLP	68
FIGURA 43. DEMOSTRACIÓN POLICY TIP	69
FIGURA 44. DEMOSTRACIÓN OVERRIDE.....	69
FIGURA 45. CONFIGURACIÓN DLP, NOTIFICACIONES EMAIL Y POLICY TIP.	70
FIGURA 46. CONFIGURACIÓN DLP, NOTIFICACIONES EMAIL Y POLICY TIP.	70
FIGURA 47. INSTALACIÓN DE AGENTE PARA VINCULAR DISPOSITIVOS	71
FIGURA 48. DISPOSITIVO 1 DADO DE ALTA EN MICROSOFT ENTRA ID	71
FIGURA 49. DISPOSITIVO 2 DADO DE ALTA EN MICROSOFT ENTRA ID	72
FIGURA 50. DISPOSITIVO 3 DADO DE ALTA EN MICROSOFT ENTRA ID	72
FIGURA 51. DISPOSITIVOS DADOS DE ALTA EN MICROSOFT PURVIEW.....	73
FIGURA 52. PANTALLA CONFIRMACIÓN INTUNE	73
FIGURA 53. VINCULACIÓN USUARIO CON DISPOSITIVO	74
FIGURA 54. IRM SETTINGS - ANONIMATO.....	75
FIGURA 55. IRM SETTINGS - WHITELIST.....	75
FIGURA 56 IRM SETTINGS – DOMINIOS ALTO RIESGO.....	76
FIGURA 57. IRM FILE UPLOADS – CONTENIDO PRIORITARIO	78

FIGURA 58. IRM FILE UPLOADS – DETECCIÓN DE TODA LA ACTIVIDAD GENERADA	78
FIGURA 59. IRM FILE UPLOADS – EVENTOS DESENCADENANTES.....	79
FIGURA 61. IRM FILE UPLOADS – UMBRALES DE EVENTOS 2.....	80
FIGURA 62. IRM FILE UPLOADS – UMBRALES DE EVENTOS 3.....	80
FIGURA 63. IRM FILE UPLOADS – UMBRALES DE EVENTOS 4.....	81
FIGURA 64. IRM FILE UPLOADS – UMBRALES DE EVENTOS 5.....	81
FIGURA 65. IRM FILE UPLOADS – INDICADORES DE RIESGO	82
FIGURA 66. IRM EXTERNAL EMAILS WITH ATTACHMENTS – EVENTOS DESENCADENANTES.....	83
FIGURA 67. IRM EXTERNAL EMAILS WITH ATTACHMENTS – UMBRALES EVENTOS DESENCADENANTES.....	84
FIGURA 68. IRM EXTERNAL EMAILS WITH ATTACHMENTS – INDICADORES DE RIESGO	84
FIGURA 69. DLM – ALMACENAMIENTO LOCAL INICIAL.....	86
FIGURA 70. EJECUCIÓN SCRIPT EXTRACT_FIELDS, PRIMER NIVEL.....	86
FIGURA 71. EJECUCIÓN SCRIPT EXTRACT_FIELDS, SEGUNDO NIVEL	86
FIGURA 72. EJECUCIÓN SCRIPT EXTRACT_FIELDS, FICHEROS	87
FIGURA 73. EJECUCIÓN SCRIPT CREATE_SITES.....	88
FIGURA 74. EJECUCIÓN SCRIPT CREATE_SITES, PRIMER NIVEL DE DIRECTORIOS.....	88
FIGURA 75. EJECUCIÓN SCRIPT CREATE_SITES, SEGUNDO NIVEL DE DIRECTORIOS.....	88
FIGURA 76. PROPIEDADES MAPEABLES	90
FIGURA 77. VÍNCULO ENTRE PROPIEDAD MAPEABLE Y PROPIEDAD REFINABLE.....	90
FIGURA 78. RESUMEN FASES VÍNCULO PROPIEDAD MAPEABLE Y PROPIEDAD REFINABLE (MICROSOFT, 2022).....	91
FIGURA 79. CREACIÓN DE QUERY DE ADAPTIVE SCOPE.....	92
FIGURA 80. ASIGNACIÓN TEMPORAL DE DOCE AÑOS – DLM	92
FIGURA 81. ASIGNACIÓN TEMPORAL DE TRES AÑOS – DLM.....	93
FIGURA 82. ASIGNACIÓN TEMPORAL DE UN AÑO – DLM.....	93
FIGURA 83. ETIQUETA DE RETENCIÓN – DLM	94
FIGURA 84. ADMINISTRADOR DEL SERVIDOR.....	101
FIGURA 85. INFORMACIÓN DEL SISTEMA	102
FIGURA 86. EQUIPO GESTIONADO POR EL DOMINIO.....	102
FIGURA 87. ACCESIBILIDAD DEL CLIENTE AL SERVIDOR	102
FIGURA 88. USUARIOS ORIGINARIOS DE CLOUD	103
FIGURA 89. USUARIOS SINCRONIZADOS	103
FIGURA 90. SINGLE SIGN-ON.....	103
FIGURA 91. EJEMPLO CONFIDENTIAL UNENCRYPTED, POLICY TIP.....	105
FIGURA 92. EJEMPLO CONFIDENTIAL UNENCRYPTED, INVALIDACIÓN	106
FIGURA 93. RECOMENDACIÓN CAMBIO DE ETIQUETA – CONFIDENTIAL ENCRYPTED.....	106
FIGURA 94. EJEMPLO ETIQUETADO - CONFIDENTIAL ENCRYPTED.....	107
FIGURA 95. ACCESO A CORREO BLOQUEADO – CONFIDENTIAL ENCRYPTED	107
FIGURA 96. ACCESO A CORREO BLOQUEADO – CONFIDENTIAL ENCRYPTED	108
FIGURA 97. ACCESO A CORREO BLOQUEADO – CONFIDENTIAL ENCRYPTED	108
FIGURA 98. APLICACIÓN A DOCUMENTO AUTOMÁTICA – INDUSTRIAL PROPERTY USE.....	109
FIGURA 99. NEGACIÓN DE ACCESO – INDUSTRIAL PROPERTY USE	109
FIGURA 100. APLICACIÓN A CORREO AUTOMÁTICA – INDUSTRIAL PROPERTY USE	109
FIGURA 101. VISTA PREVIA ANÓNIMA – FILE UPLOADS.....	111
FIGURA 102. INTRODUCCIÓN DE PROMPT – FILE UPLOADS.....	111
FIGURA 103. SOLICITAR INFORMACIÓN SENSIBLE – FILE UPLOADS.....	112
FIGURA 104. RESUMEN DE EVENTOS COPILOT – FILE UPLOADS	112
FIGURA 105. LISTADO DE ALERTAS “RISKY PROMPT ENTERED” – FILE UPLOADS.....	113
FIGURA 106. DETALLE DE EVENTO “RISKY PROMPT ENTERED” – FILE UPLOADS	113
FIGURA 107. DETALLE DE LA INFORMACIÓN EN EL EVENTO “RISKY PROMPT ENTERED” – FILE UPLOADS	114
FIGURA 108. PROMPT DE RIESGO DEL EVENTO “RISKY PROMPT ENTERED” – FILE UPLOADS.....	114
FIGURA 109. DETECCIÓN DE INFORMACIÓN SENSIBLE INVENTION PROPOSAL– FILE UPLOADS.....	115
FIGURA 110. DETECCIÓN DE ETIQUETADO SENSIBLE INDUSTRIAL PROPERTY USE – FILE UPLOADS	115
FIGURA 111. EVENTO SENSITIVE RESPONSE COPILOT – FILE UPLOADS	116
FIGURA 112. LISTADO DE EVENTOS SENSITIVE RESPONSE COPILOT – FILE UPLOADS.....	116
FIGURA 113. DETALLE DE ACTIVIDAD SENSITIVE RESPONSE COPILOT – FILE UPLOADS.....	116
FIGURA 114. DETALLE DE INFORMACIÓN SENSIBLE SENSITIVE RESPONSE COPILOT – FILE UPLOADS.....	117
FIGURA 115. DETALLE ACTIVIDAD SENSIBLE INDUSTRIAL PROPERTY USE, RESPONSE COPILOT – FILE UPLOADS	117
FIGURA 116. DETALLE DEL PROMPT, SENSITIVE RESPONSE COPILOT – FILE UPLOADS.....	117
FIGURA 117. DETALLE DE ACTIVIDAD USUARIO, SUBIDAS DOMINIOS WEB– FILE UPLOADS	118

FIGURA 118. LISTADO EVENTOS GENERADOS, SUBIDAS DOMINIOS WEB – FILE UPLOADS 118

FIGURA 121. RESUMEN DE ALERTA SIN ANONIMIZAR – FILE UPLOADS..... 120

FIGURA 122. DETALLE DE EVENTOS GENERADOS, ALERTA SIN ANONIMIZAR – FILE UPLOADS 120

FIGURA 123. LISTADO DE EVENTOS GENERADOS, ALERTA SIN ANONIMIZAR – FILE UPLOADS 120

FIGURA 124. DETALLE DE ACTIVIDAD, ALERTA SIN ANONIMIZAR – FILE UPLOADS..... 121

FIGURA 125. DETALLE DE LA ALERTA EXTERNAL EMAILS WITH ATTACHMENTS..... 122

FIGURA 126. DETALLE DE ACTIVIDAD DEL USUARIO - EXTERNAL EMAILS WITH ATTACHMENTS..... 122

FIGURA 127. LISTADO DE EVENTOS DE LA ALERTA - EXTERNAL EMAILS WITH ATTACHMENTS 122

FIGURA 128. DETALLES DE LA ACTIVIDAD - EXTERNAL EMAILS WITH ATTACHMENTS..... 123

FIGURA 129. DETALLES DEL ÍTEM DE LA ACTIVIDAD- EXTERNAL EMAILS WITH ATTACHMENTS..... 123

FIGURA 130. RESUMEN DE LOS DOCUMENTOS ADJUNTOS - EXTERNAL EMAILS WITH ATTACHMENTS 124

FIGURA 131. RESUMEN DE LOS DOCUMENTOS ADJUNTOS - EXTERNAL EMAILS WITH ATTACHMENTS 124

FIGURA 132. APLICACIÓN DE PROPIEDAD MAPEABLE – DLM..... 125

FIGURA 133. LISTADO DE SITIOS DE SHAREPOINT CONSULTADOS – DLM..... 125

FIGURA 134. LISTADO DE SITIOS DE SHAREPOINT, ÁMBITO DINÁMICO– DLM 126

FIGURA 135. LISTADO DE POLÍTICAS APLICADAS A SITIOS DE SHAREPOINT– DLM 126

1 Introducción

Hoy en día la protección de los datos en una organización es crucial para garantizar la seguridad de estas. La utilización de tecnologías para protegerlos permite a las empresas asegurar la confidencialidad, integridad y disponibilidad de la información, así como cumplir con normativas legales y evitar pérdidas económicas o daños reputacionales. En entornos como los despachos de abogados, donde se manejan datos especialmente sensibles, resulta imprescindible disponer de mecanismos avanzados de control y protección que aseguren el tratamiento adecuado de dicha información. Es por ello por lo que este trabajo de fin de grado tiene como objetivo desarrollar una solución integral basada en la nube de Azure para proteger y gestionar la información de un despacho ficticio de abogados de patentes y marcas.

La solución presentada se basa en el desarrollo de tres fases diferenciadas, teniendo por una parte el desarrollo de una solución para la gobernanza de la identidad, siendo esta la proporcionada por configurar una arquitectura de gestión de usuarios híbridos. Mediante esta solución, el administrador puede gestionar roles y permisos de un usuario tanto de forma local como de forma en la nube. Además, gracias a esta implementación el usuario es capaz de acceder tanto a su entorno local como a su entorno en la nube con un solo inicio de sesión y con solo un tipo de credencial, todo vinculado a su dispositivo de trabajo. Todo esto permite a una empresa migrar progresivamente su entorno local si así lo desea, o mantener ambos enfoques.

Por otra parte, se implementó todo lo relacionado a la protección de los datos utilizando tecnologías Microsoft Azure Purview:

- **Microsoft Information Protection:** con esta tecnología obtenemos la creación de directivas para clasificar la información según el contenido de los documentos. Esto nos permite limitar accesos mediante la asignación de permisos y encriptación.
- **Data Loss Prevention:** con esta funcionalidad se definió una política de prevención de pérdida de datos encargada de detectar si un usuario está enviando información al exterior, solicitando un motivo para ello mediante la aparición de un policy tip (pop-up).
- **Data Lifecycle Management:** con esta tecnología definimos políticas de retención para retener sitios de SharePoint en base a ámbitos dinámicos (Adaptive Scopes) de pares de atributos determinados por clave-valor.
- **Insider Risk Management:** se definieron políticas para controlar la pérdida de datos derivada a actividad interna en la empresa. Se evalúa si un usuario exfiltra documentación al exterior.

Por último, se desarrolló una aplicación web para la gestión de citas. Los clientes pueden reservar una franja horaria disponible para un trámite específico. A partir de esta reserva, se generan automáticamente un acuerdo de confidencialidad y una factura, los cuales se almacenan en el SharePoint del inquilino (tenant) de Azure de la empresa gracias a la integración con la API de Microsoft Graph. Estos documentos son clasificados y se les aplican las políticas de retención correspondientes al sitio de almacenamiento. Además, la cita se agenda en el calendario del empleado asignado y se envían correos de confirmación tanto al empleado como al cliente. Toda esta solución fue implementada mediante una API REST desarrollada con Spring Boot, un frontend construido con React y TypeScript, y el uso integrado de la API de Microsoft Graph.

2 Palabras clave del proyecto

Los términos que describen mejor el proyecto son los siguientes:

- Seguridad y privacidad (99999912)
- Programación web (99999907)
- Programación (77011012)
- Informática aplicada (77011001)
- Seguridad Informática (77011005)
- Ingeniería del software (77011007)
- Lenguajes de programación (77011012)
- Redes (99999937)

3 Descripción general del proyecto

Como se mencionó en la introducción, este proyecto tiene como objetivos, por un lado, implementar una infraestructura que permita a una empresa gestionar de forma segura la identidad híbrida, así como la protección y el ciclo de vida de sus datos. Por otro lado, también se persigue la automatización del proceso de reserva de citas, así como la generación de facturas y acuerdos de confidencialidad mediante una aplicación web.

Cada empresa es diferente, por lo tanto, tienen diferentes necesidades y previsiones de uso, en base a ello se implementa un entorno que las solventa. A continuación, se detalla cada punto.

3.1 Necesidades

El caso de uso que se ha utilizado en este trabajo de fin de grado viene dado por un despacho ficticio de abogados de patentes, diseños industriales, modelos de utilidad y marcas. Se han planteado las siguientes necesidades del cliente:

3.1.1 Organización de documentos locales

El cliente tiene todos sus documentos de registros de ideas relativos a la propiedad intelectual almacenados de forma local, en directorios de su propio sistema de ficheros. Los documentos del caso de uso son registros de patentes, marcas y modelos de utilidad en España, registros de diseños industriales en la Unión Europea y registros de marcas en Estados Unidos. Se ha previsto que estos directorios están desorganizados y mezclados, teniendo en un mismo directorio, mezclas entre trámites con una muy mala gestión y control de su información.

De este punto surge la necesidad de idear un sistema escalable y mantenible en el tiempo, con un acceso fácil y práctico. Estas características nos la proporciona el utilizar almacenamiento en la nube, que en este caso sería con Azure, y utilizar su herramienta de espacios de trabajo compartidos SharePoint. A este espacio de trabajos compartidos tendrá acceso todos los abogados de la firma con acceso permitido, lo cual facilita el orden, protección y gestión de los datos.

Para poder automatizar la migración de estos archivos locales a SharePoint, lo que se ha hecho es idear dos scripts:

- Extract_fields.py: script en Python que se encarga de recorrer el directorio local desorganizado, lee cada fichero pdf y según el país, autor y tipo de registro genera una estructura de directorios que luego será la estructura de sites de SharePoint. Asimismo, controla que no se duplique ningún directorio o fichero.
- Create_Sites.ps1: script en PowerShell usando el módulo de PnP.Online el cual se encarga de recorrer el árbol de directorios generado con el script anterior y crea los sites de SharePoint correspondientes a cada directorio padre (País_Trámite) y un subdirectorio para cada cliente para almacenar dentro los documentos de registro realizados para ese cliente. Este Script controla la gestión de duplicados, antes de generar un sitio o directorio verifica si existe, y si no existe, lo crea.

En el apartado de diseño se explicará toda la estructura de los Sites de SharePoint y en la sección de implementación se detallará todo el flujo de creación y el resultado obtenido.

3.1.2 Cumplimiento de periodos de retención de los documentos

Según el tipo de trámite registrado, habrá un periodo de renovación más o menos extenso. Cuando hay una gran cantidad de registros y flujos de información, resulta conveniente tener políticas que retengan estos datos o los eliminen, es por ello por lo que se implementaron políticas de retención mediante la herramienta Data Lifecycle Management de Microsoft Purview con ámbitos adaptativos a cada Site de SharePoint según su tiempo de retención.

Gracias a esta implementación se cumple la necesidad de ahorro de espacio, ya que definir que un documento se retenga hasta un cierto periodo de tiempo nos permite la eliminación automática de documentos que no tienen ningún uso y solo están ocupando espacio en la empresa. Por ejemplo, el registro de un modelo de utilidad no necesita el mismo nivel de mantenimiento que el registro de una patente, en base a esto se retiene en la base de datos un documento más o menos tiempo.

- `config_sites.ps1`: para poder cumplir con el cometido de configurar políticas de retención y que sea aplicado cada tipo a cada site, se desarrolló este script en PowerShell.

En el apartado de diseño e implementación de Data Lifecycle Management se explicará en detalle este script y su funcionamiento.

3.1.3 Protección de la información confidencial

En la explicación del caso de uso se plantea la necesidad de mantener protegida la información que la empresa considera sensible. Esta información abarca los siguientes temas:

- Información referente a acuerdos de confidencialidad.
- Información personal de los usuarios, como DNI o tarjetas de débito/crédito.
- Información referente a registros pertenecientes al campo de la propiedad intelectual.
- Información financiera referente a facturas.

Esta necesidad de proteger la información será solventada mediante la utilización de los siguientes mecanismos:

1. Etiquetado de la información mediante etiquetas configuradas en Microsoft Information Protection. Esto permite limitar los permisos de acceso a un documento.
2. Definición de políticas de detección de riesgos internos mediante Insider Risk Management para detectar cuándo un empleado realiza una acción de exfiltración de información sensible de la empresa.

3.1.4 Migración de usuarios e inicio de sesión único

Se ha planteado que la empresa tiene la necesidad de volver sus sistemas más escalables, seguros, accesibles desde cualquier ubicación y con una menor dependencia de la infraestructura local, mediante la centralización del contenido de sus controladores de dominios de Active Directory al cloud de Azure.

Asimismo, hacer esta migración tiene múltiples ventajas, algunas de ellas son:

- Reducción de costes operativos
- Mayor resiliencia ante desastres
- Mejora en la disponibilidad del servicio
- Simplificación de la administración de identidades

- Acceso a funcionalidades avanzadas de seguridad como el inicio de sesión sin contraseña, el acceso condicional y la autenticación multifactorial.
- Acceso remoto seguro a los recursos sin necesidad de una VPN

Una de las finalidades en este trabajo de fin de grado es demostrar que es posible realizar una migración progresiva al cloud, pudiendo mantener inicialmente una identidad híbrida que permitirá a los empleados iniciar sesión con las mismas credenciales y una sola vez tanto en su entorno local como en su entorno en la nube.

Esta creación de identidad híbrida ha sido posible gracias a la utilización de Microsoft Entra Connect, proceso el cual será explicado más adelante en este trabajo de fin de grado.

3.1.5 Monitorización de riesgos internos

La empresa tiene la necesidad de monitorizar la actividad de sus empleados para garantizar que no se está incumpliendo con las políticas de protección de sus datos. Concretamente, requiere monitorizar en dos casos específicos:

- Subidas a dominios web: toda información proveniente de documentos adjuntos de la empresa subida a dominios web debe ser controlada y monitorizada. Para ello, se creó la política de IRM File Uploads.
- Interacciones de riesgo con Copilot: una nueva necesidad de los tiempos modernos es proteger las interacciones con Copilot para prevenir fugas de información sensible y posible compartición con otro individuo. Para monitorizar esta actividad se incluyó en la política de IRM File Uploads una sección que monitoriza estos eventos.
- Compartición de archivos con externos: Se tiene la necesidad de monitorizar los documentos que se comparten con usuarios externos, por ello, se implementó la política de IRM External emails with attachments.

3.1.6 Gestión de citas de forma automatizada

Otra necesidad de la firma de abogados ficticia es la de automatizar su proceso de gestión de citas, de manera que un cliente pueda ingresar en una aplicación web y agendar una cita para el trámite que necesite, según las horas disponibles de los empleados, en tiempo real.

Esta necesidad se resuelve mediante la implementación de una página web capaz de consultar la disponibilidad de los empleados, agendar una cita con ellos, generar acuerdos de confidencialidad en función del trámite a realizar y emitir las facturas correspondientes. Todos estos documentos se suben automáticamente al SharePoint destinado a las citas pendientes, y se envía una notificación por correo electrónico tanto al cliente como al empleado involucrado. Todo ello es posible gracias a la integración de la API desarrollada con la API de Microsoft Graph.

3.2 Previsiones de uso

3.2.1 Uso de scripts para los documentos locales

Para el directorio local desorganizado explicado anteriormente, se realizará la ejecución de los scripts explicados anteriormente para obtener la estructura de Sites en SharePoint. Como controla la creación de duplicados, no se generará ningún error si se ejecuta dos veces bajo el mismo directorio padre.

3.2.2 Uso de scripts para el cumplimiento del periodo de retención

Se hará uso del script config_sites.ps1 cada vez que se necesite asignar un nuevo periodo de retención a un site de SharePoint.

3.2.3 Uso de etiquetas de sensibilidad

Se hará uso de las etiquetas de sensibilidad configuradas según mecanismos de los siguientes tipos:

1. **Recomendación:** El sistema es capaz de detectar cuando se introduce información confidencial que debería estar etiquetada y le recomienda al usuario la aplicación de la etiqueta. Esto se hace con la finalidad de generar consciencia en el empleado sobre la confidencialidad de información que maneja. Es una estrategia de concientización que prevé a largo plazo la protección de la información voluntaria de cada empleado a medida que aprenda lo que debe hacer para ello.
2. **Automatización:** En el caso de que se genere un acuerdo de confidencialidad o una factura, el sistema automáticamente aplicará una etiqueta de sensibilidad para proteger la información.
3. **Sobreescritura (Override):** cuando un usuario intente enviar información desde su dirección de correo electrónico a un externo, aparecerá un aviso que pide la justificación de dicho envío. Esto es debido a que toda la información compartida desde el puesto de trabajo se considera privada y requiere justificación.

3.2.4 Inicio de sesión único para los empleados

Al estar implementada la identidad híbrida, los usuarios podrán autenticarse una sola vez con sus credenciales corporativas para acceder tanto a recursos locales como a servicios en la nube sin necesidad de múltiples inicios de sesión.

3.2.5 Gestión centralizada para los administradores del sistema

El administrador del sistema ahora podrá añadir nuevos usuarios y grupos y administrarlos en la nube, pero, además, también podrá administrar todo lo que había creado previamente de forma local. Esto dará una previsión de uso de identidades unificadas, con todos los beneficios que esto conlleva.

3.2.6 Reserva de citas y generación de documentos

Los clientes son capaces de reservar citas con los empleados, en una franja de hora, en el servicio que necesiten.

Además, mediante la cumplimentación de un formulario de acuerdo de confidencialidad en el flujo de reserva de cita, el proceso de firma de estos documentos para proteger las ideas no registradas estará protegida de forma automatizada.

3.2.7 Revisión de alertas de riesgos internos

El administrador del sistema es capaz de revisar las alertas generadas por los usuarios de la empresa en Insider Risk Management gracias a la implementación de las tecnologías File Uploads y External emails with attachments.

4 Objetivos

A pesar de que puede haberse explicado anteriormente de forma indirecta los objetivos de este trabajo de fin de grado, a continuación, se presenta explícitamente cada uno de ellos, y su justificación formativa.

4.1 Objetivo general

El objetivo general de este Trabajo de Fin de Grado es el diseño e implementación de una solución integral para la protección y gestión de la información en una empresa ficticia de un despacho de abogados especializado en la propiedad industrial, utilizando herramientas del ecosistema Microsoft Purview, Microsoft Entra ID, Microsoft 365 y tecnologías web modernas como SpringBoot y React. Esta solución busca cubrir desde la gestión de la identidad híbrida hasta la clasificación automática de documentos, la creación de políticas de retención y la integración de una aplicación web que interactúe con dicha infraestructura y le dé un uso global a la solución propuesta.

4.2 Objetivos específicos

A partir del objetivo general, se establecieron los siguientes objetivos específicos:

- **Diseñar una infraestructura de identidad híbrida:** mediante la integración de Active Directory local con Microsoft Entra Connect, permitiendo un inicio de sesión único (SSO) y sincronización de usuarios y grupos entre entornos locales y en la nube.
- **Definir y aplicar estrategias de protección de la información:** se van a utilizar la mayoría de las herramientas de las que se dispone en Microsoft Purview.
 1. Data Loss Prevention (DLP): Para la creación de una política que permita sobrescribir una política de etiqueta de la empresa para permitir al usuario proporcionar una justificación.
 2. Microsoft Information Protection (MIP): Usada para la clasificación de documentos aplicando estrategias de recomendación y automatización, además de la forma manual tradicional.
 3. Insider Risk Management (IRM): Para crear dos políticas que monitoricen actividad de riesgo interna en la empresa. Se pretende monitorizar subidas a dominios externos, compartición de información sensible con Copilot y el envío de correos con archivos adjuntos a personas externas a la organización.
- **Automatizar la creación de entornos colaborativos en SharePoint:** derivado de la creación de estos entornos, se pretende evitar el almacenamiento local desorganizado, además de asegurar los periodos de retención que la empresa necesita, aplicando políticas de retención con la herramienta de Purview Data Lifecycle Management (DLM), basadas en metadatos personalizados mediante scripts en PowerShell.
- **Desarrollar una aplicación web funcional,** con backend en Spring Boot y frontend en React con TypeScript, que permita gestionar citas legales, generar documentos de confidencialidad y facturas, generar una notificación de confirmación por correo electrónico de la cita y registrar eventos en calendarios de Microsoft 365.

4.3 Justificación de la elección del tema

La elección de este tema se basa en la constante preocupación del sector tecnológico en proteger la información. En un mundo donde la tecnología forma parte de la vida cotidiana y los datos pueden ocasionar pérdidas económicas, legales o reputacionales inmensas, es imprescindible que los profesionales de la informática adquieran competencias sólidas en materia de ciberseguridad, gobierno de la información y gestión en la nube.

Desde el punto de vista formativo, este trabajo ha supuesto una oportunidad para consolidar y aplicar múltiples conocimientos adquiridos durante el grado: desde la administración de sistemas, redes y automatización con scripts, hasta el desarrollo web orientado a modelo-vista-controlador con una API Rest y desarrollo de frontend utilizando una tecnología con diversos patrones de diseño.

Así mismo, desde el punto de vista formativo, además de poder poner en práctica las habilidades y conocimientos adquiridos a lo largo de toda la carrera, también se ha hecho una importante fase de investigación de tecnologías Microsoft Purview, conllevando un autoaprendizaje constante. La elección de estos temas me permite desarrollarme en términos de protección de datos en tecnologías referente a nivel mundial como lo son las tecnologías de Microsoft.

La simulación de un entorno profesional como el de un despacho de abogados ha permitido contextualizar de forma realista los retos actuales en torno a la protección de datos sensibles, fomentando no solo habilidades técnicas sino también una visión crítica y responsable de los propios empleados sobre el uso de la tecnología en entornos empresariales.

5 Planificación

Para poder desarrollar este trabajo de fin de grado se ha desarrollado un calendario desde el día 10 de febrero hasta el día 6 de junio. Al ser un TFG con fases diferenciadas, se ha organizado el tiempo en base a ello, dando como resultado el siguiente diagrama de Gantt:

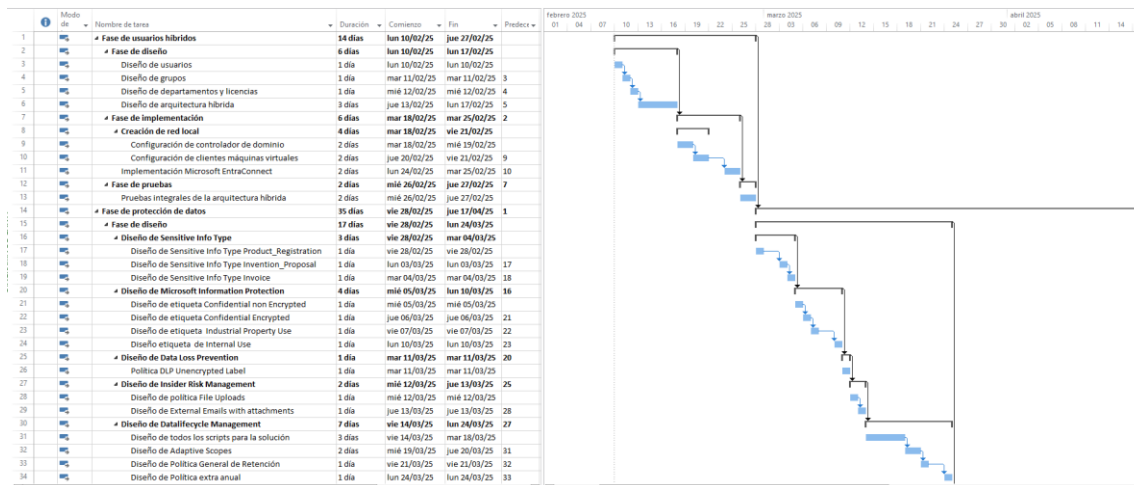


Figura 1. Diagrama de Gantt, tareas 1-34

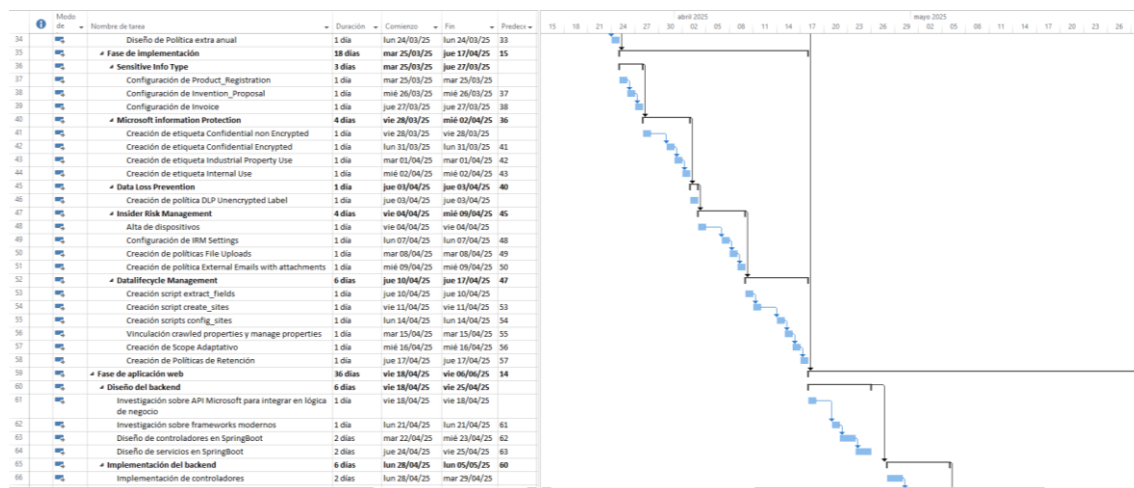


Figura 2. Diagrama de Gantt, tareas 34-66

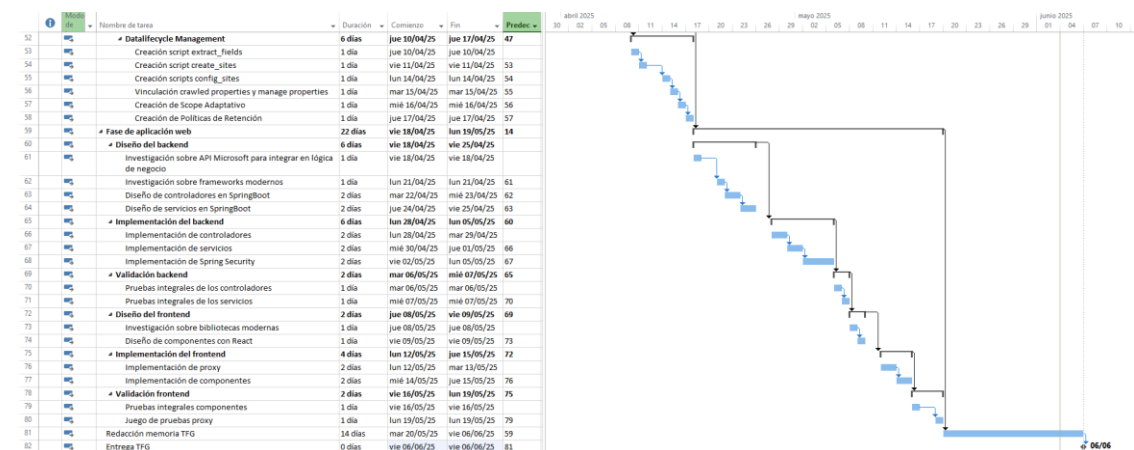


Figura 3. Diagrama de Gantt, tareas 66-82

6 Requisitos

A continuación, se plantean los requisitos funcionales y no funcionales.

6.1 Requisitos Funcionales

6.1.1 Fase identidad híbrida

Esta fase comprende la creación de la identidad híbrida con Microsoft Entra Connect.

6.1.1.1 Reglas del negocio

01. No puede haber dos usuarios con el mismo UPN (User Principal Name, (tiene formato de correo electrónico) en el entorno híbrido. El UPN debe ser único en Microsoft Entra ID y local Active Directory.
02. Cada cuenta local debe estar sincronizada una sola vez. No se permite que una misma cuenta de Active Directory esté duplicada en Microsoft Entra ID.
03. No se puede sincronizar un usuario sin un atributo de UPN válido, debe seguir las convenciones de la empresa.
04. El atributo sAMAccountName (nombre de usuario) debe ser único en el dominio local.
05. No se puede habilitar el inicio de sesión en la nube para cuentas sin contraseña o bloqueadas en el entorno local.
06. No se puede instalar Microsoft Entra Connect en un equipo que no sea controlador de dominio.
07. Cada grupo sincronizado debe cumplir con las restricciones de tamaño de Microsoft Entra ID.

6.1.1.2 Especificación textual

A continuación, se explicará toda la especificación textual de la fase de usuarios híbridos.

6.1.1.2.1 Casos de uso

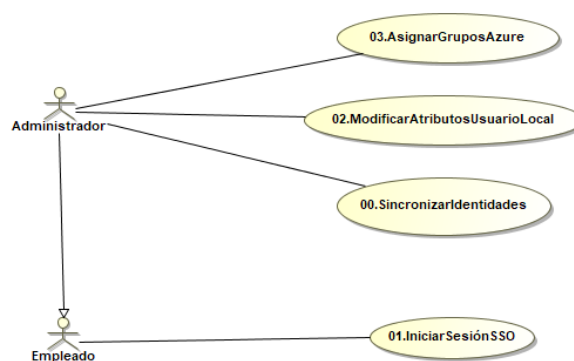


Figura 4. Diagrama casos de uso fase identidad híbrida

Caso de uso 00. SincronizarIdentities

Resumen de la funcionalidad: Permite sincronizar usuarios del Active Directory local a Microsoft Entra ID con el uso de Entra ID Connect.

Parámetro de entrada: Ninguno.

Parámetro de salida: Ninguno.

Actores: **Administrador del sistema.**

Precondición: Entra ID Connect debe estar instalado y configurado correctamente.

Postcondición: Los usuarios del entorno local se encuentran disponibles en Microsoft Entra ID.

Proceso normal principal:

1. El **administrador** ejecuta el asistente de sincronización de Entra Connect o espera a que se ejecute la sincronización automática.
2. El sistema comprueba la conexión con el controlador de dominio local y con Microsoft Entra ID.
3. El sistema recopila los objetos de usuarios desde las unidades organizativas configuradas.
4. El sistema valida que los atributos necesarios como UPN, mail, etc, sean válidos.
5. El sistema sincroniza los objetos con Microsoft Entra ID.
6. El sistema genera un informe

Alternativas de proceso y excepciones:

5a. Si un usuario ya existe en Microsoft Entra ID con los mismos atributos, el sistema marca conflicto y lo notifica en el informe.

5b. Si hay errores en la conectividad o replicación del dominio, se detiene la sincronización.

Caso de uso 01. IniciarSesionSSO

Resumen de la funcionalidad: Permite que un usuario inicie sesión una sola vez para acceder tanto a recursos locales como a recursos en la nube.

Parámetro de entrada: Credenciales del usuario.

Parámetro de salida: Inicio de sesión

Actores: **Empleado.**

Precondición: El usuario debe estar sincronizado en Microsoft Entra ID.

Postcondición: El usuario puede acceder a recursos tanto en local como en la nube con una sola autenticación.

Proceso normal principal:

1. El usuario introduce sus credenciales en el inicio de sesión de su ordenador.
2. El sistema valida las credenciales en el Active Directory local.
3. Se emite un inicio de sesión válido para el entorno local.
4. El sistema proporciona acceso también a recursos cloud como Microsoft 365, SharePoint, Exchange, sin requerir nuevo inicio de sesión.
5. El usuario accede a recursos de forma transparente.

Alternativas de proceso y excepciones:

- 2a. Si las credenciales no son válidas, se deniega el acceso y se muestra mensaje de error.
- 4a. Si hay pérdida de conectividad con Azure, el sistema permite únicamente acceso local a los recursos.

Caso de uso 02. ModificarAtributosUsuarioLocal

Resumen de la funcionalidad: Permite modificar atributos de una cuenta de usuario en el Active Directory local, que luego serán sincronizados con Microsoft Entra ID.

Parámetro de entrada: Identificador del usuario, atributos a modificar.

Parámetro de salida: Ninguno.

Actores: **Administrador del sistema.**

Precondición: El usuario debe existir en el entorno local.

Postcondición: Los cambios serán replicados en Microsoft Entra ID tras la próxima sincronización.

Proceso normal principal:

1. El **administrador** edita el perfil del usuario en Active Directory (por ejemplo, cambia el departamento).
2. El sistema almacena los cambios localmente.
3. Entra ID Connect detecta el cambio durante la siguiente sincronización.
4. El atributo modificado se actualiza también en Microsoft Entra ID.

Alternativas de proceso y excepciones:

- 3a. Si el nuevo valor del atributo entra en conflicto con otro objeto sincronizado, se genera un error en el informe de sincronización.

Caso de uso 03. AsignarPermisosGrupoAzure

Resumen de la funcionalidad: Permite asignar a un usuario permisos sobre recursos cloud mediante su incorporación a grupos en Microsoft Entra ID.

Parámetro de entrada: Identificador del usuario, identificador del grupo.

Parámetro de salida: Ninguno.

Actores: **Administrador del sistema.**

Precondición: El usuario debe estar sincronizado en Microsoft Entra ID.

Postcondición: El usuario tiene acceso a los recursos asociados al grupo asignado.

Proceso normal principal:

1. El **Administrador** accede al portal de Entra ID.
2. Selecciona el grupo de seguridad correspondiente.
3. Añade al usuario deseado al grupo.
4. El sistema actualiza la membresía del grupo.
5. Azure aplica automáticamente los permisos asociados al grupo sobre los recursos correspondientes.

Alternativas de proceso y excepciones:

3a. Si el grupo es dinámico y el usuario no cumple las condiciones, el sistema no permite su incorporación.

6.1.2 Fase protección de datos

En esta fase se especifica todos los requisitos funcionales de la fase de protección de datos. Se ha decidido separarlo en fases debido a que era la forma más ordenada de mostrar la información.

6.1.2.1 Reglas del negocio

6.1.2.1.1 Microsoft Information Protection

01. Todo documento o correo debe tener una etiqueta de sensibilidad aplicada.
02. La etiqueta "Internal Use" debe aplicarse por defecto a todo documento o correo que no contenga información sensible. Es la etiqueta por defecto de la empresa.
03. Si un documento contiene información sensible del tipo "Product_Registration" o "Invention_Proposal", debe aplicarse automáticamente la etiqueta "Industrial Property Use".
04. Solo los usuarios del grupo "Intellectual Property Agent" pueden acceder a contenido con la etiqueta "Industrial Property Use".
05. Los documentos con información del tipo "Invoice", "DNI", "IBAN", "tarjetas bancarias", "Número de pasaporte", deben recibir una recomendación automática para aplicar la etiqueta "Confidential Encrypted".
06. La etiqueta "Confidential Encrypted" debe cifrar la información de los documentos o correos que se compartan.
07. La etiqueta "Confidential Encrypted" solo permite el acceso a los empleados registrados en el tenant de la empresa.
08. La etiqueta "Confidential non Encrypted" no debe cifrar la información de los documentos o correos que se compartan.
09. La etiqueta "Confidential non Encrypted" solo puede aplicarse manualmente, y su uso conlleva una alerta y requerimiento de justificación si el contenido se envía a un externo.
10. No se puede enviar contenido con la etiqueta "Confidential non Encrypted" a dominios externos sin justificación de la acción.

6.1.2.1.2 Sensitive Info Type

01. El Sensitive Info Type "Invoice" debe detectar los patrones de divisas tales como €50, 50€, \$100 y estos deben ser identificados como información financiera sensible.

02. El Sensitive Info Type “Invention_Proposal” debe detectar por una parte contratos de confidencialidad en los idiomas español e inglés y por otra parte debe detectar propuestas de invención.
03. El Sensitive Info Type “Product_Registration” debe detectar documentos de registros de marcas en España, Estados Unidos y Unión Europea, solicitudes de modelos de utilidad en España, registros de diseños industriales en la Unión Europea y registros de patentes en Estados Unidos.

6.1.2.1.3 Insider Risk Management

01. El sistema debe generar alertas si un usuario sube archivos a dominios web no permitidos bajo la política File Uploads.
02. El sistema detectará si se envían prompts con datos sensibles a Copilot, generando una alerta bajo la política File Uploads.
03. El sistema detectará si se reciben respuestas con información sensible desde Copilot, generando una alerta bajo la política File Uploads.
04. La actividad de subir archivos a herramientas como WhatsApp Web o Google Drive debe ser monitoreada bajo la política File Uploads.
05. Todos los dispositivos que se monitorean deben estar registrados en Microsoft Purview.
06. Se debe detectar cuando se envíe un correo a un externo con adjuntos mediante la política External emails with attachments.
07. El análisis de alertas debe realizarse inicialmente sin anonimización del usuario.
08. Los dominios de confianza deben estar definidos para evitar falsos positivos.
09. Los dominios de alto riesgo deben estar definidos para priorizar las alertas.
10. Se debe priorizar las alertas generadas que contengan Sensitive Info Types.

6.1.2.1.4 Datalifecycle Management

6.1.2.1.4.1 Organización de Sites y bibliotecas

01. Cada combinación País_Trámite debe tener un Site de SharePoint independiente para organizar los documentos relacionados con sus clientes y registros.
02. Cada cliente dentro de un Site debe tener su propia carpeta con subcarpetas para facturas y registros.
03. Los empleados deben usar las bibliotecas “Clients” y “Pending Appointments” en cada Site generado para la gestión de clientes y citas.
04. No se debe crear un Site duplicado si este ya existe en el tenant; si existe, solo se actualizan los documentos.

05. La estructura de Sites debe generarse de forma automática a partir del árbol de directorios local definido mediante **extract_fields.py**.

06. El script create_sites.ps1 debe encargarse de crear los Sites y bibliotecas según el CSV generado por el script **extract_fields.py**.

6.1.2.1.4.2 *Propiedades mapeables y configuraciones por Site*

07. Cada Site debe tener al menos una propiedad refinable (RefinableString) que determine su política de retención.

08. Las propiedades mapeables deben asignarse mediante el script **config_sites.ps1** utilizando el CSV correspondiente.

09. Todos los Sites deben tener como mínimo una propiedad “GeneralRetention” para aplicar políticas globales de retención.

6.1.2.1.4.3 *Periodos de retención y Adaptive Scopes*

10. Cada Site debe tener al menos una propiedad mapeable (RefinableString) que determine su política de retención.

Se deben definir períodos de retención específicos por tipo de trámite legal y país:

- Patente o modelo de utilidad (España): 3 años
- Marca (España, USA, UE): 10 años
- Diseño industrial (UE): 5 años

11. El período de retención debe ser calculado como: frecuencia_renovación + 2 años, siempre y cuando no se haya modificado el documento.

12. El sistema debe crear un Adaptive Scope por cada valor de GeneralRetention, filtrando los Sites mediante el RefinableString00.

13. Las políticas de retención deben asociarse a los scopes adaptativos y respetar el tiempo de retención definido por tipo de documento.

14. La política Plus1Year puede aplicarse manualmente a documentos puntuales para extender su retención un año adicional.

6.1.2.1.4.4 *Control de errores*

15. Los scripts de automatización deben evitar duplicidades en la creación de Sites, bibliotecas o propiedades.

16. Cualquier Site con errores de nombre o claves mapeables inválidas debe registrarse en el log de ejecución.

6.1.2.2 Especificación textual

A continuación, se presentan los casos de uso de la fase de protección de datos. Se ha dividido en subapartados para una mejor comprensión de los procesos y sus relaciones entre soluciones.

6.1.2.2.1 Microsoft Information Protection

Esta sección comprende los casos de uso de la solución implementada con Microsoft Information Protection.

6.1.2.2.1.1 Casos de uso

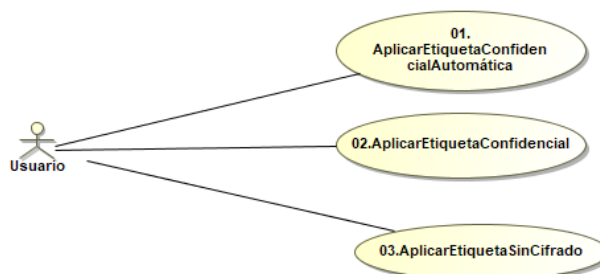


Figura 5. Diagrama de casos de uso Microsoft Information Protection

Un punto importante por aclarar antes de la explicación de los casos de uso de Microsoft Information Protection es que la estrecha relación entre la etiqueta “Confidential non Encrypted” y la política de Data Loss Prevention, ha ocasionado que estos dos puntos se fusionen y se generen un solo caso de uso, el cual es el Caso de uso 03. AplicarEtiquetaSinCifrado.

Caso de uso 01. AplicarEtiquetaConfidencialAutomatica

Resumen de la funcionalidad: El sistema aplica automáticamente una etiqueta de sensibilidad a un documento o correo si se detecta información sensible previamente definida.

Parámetro de entrada: Documento o correo redactado.

Parámetro de salida: Etiqueta de sensibilidad aplicada.

Actores: **Usuario**.

Precondición: El contenido del documento o mail debe contener información con alguno de los patrones definidos en el Sensitive Info Types de “Industrial Property Use”.

Postcondición: El contenido queda etiquetado y protegido automáticamente.

Proceso normal principal:

1. El **Usuario** redacta un documento o correo.
2. El sistema analiza el contenido mediante los patrones definidos en Microsoft Purview.
3. Si detecta coincidencia con el tipo de información sensible “Invention_Proposal”, aplica automáticamente la etiqueta correspondiente
4. Se limitan los permisos de acceso según la configuración de la etiqueta.
5. Se registra el evento en el Activity Explorer.

Alternativas y excepciones:

3a. Si el contenido no coincide con ninguna definición sensible, se aplica la etiqueta por defecto "Internal Use".

Caso de uso 02. AplicarEtiquetaConfidencial

Resumen de la funcionalidad: Se sugiere al usuario aplicar una etiqueta cifrada si se detecta contenido sensible relativo a datos financieros o personales.

Parámetro de entrada: Documento o correo redactado.

Parámetro de salida: Recomendación de cambio de etiqueta.

Actores: **Usuario**.

Precondición: El contenido debe contener datos sensibles como DNI, IBAN, tarjetas, o patrones de divisa.

Postcondición: El contenido es cifrado si el usuario acepta la recomendación.

Proceso normal principal:

1. El **Usuario** redacta un correo/documento
2. El sistema detecta información sensible según el tipo "Invoice" o campos personales.
3. El sistema muestra un Policy Tip recomendando aplicar la etiqueta "Confidential Encrypted".
4. El **Usuario** acepta la sugerencia.
5. El contenido se cifra automáticamente al aplicar la etiqueta de confidencialidad recomendada.

Alternativas y excepciones:

4a. El **usuario** ignora la recomendación, y el contenido queda con la etiqueta anterior. Esto se registra en el Activity Explorer de Purview para revisión de cumplimiento.

Caso de uso 03. AplicarEtiquetaSinCifrado

Resumen de la funcionalidad: Evita que un usuario envíe contenido no cifrado a un destinatario externo si no justifica su decisión.

Parámetro de entrada: Correo con contenido etiquetado como "Confidential non Encrypted".

Parámetro de salida: Justificación o bloqueo del envío.

Actores: **Usuario**

Precondición: El destinatario no pertenece a la organización y se aplica la etiqueta "Confidential non Encrypted".

Postcondición: El envío se bloquea o se permite tras justificación.

Proceso normal principal:

1. El **Usuario** aplica la etiqueta "Confidential non Encrypted" en un correo.
2. El **Usuario** añade un destinatario externo.
3. El sistema detecta que se intenta enviar un correo no cifrado a un destinatario externo.
4. El sistema muestra un Policy Tip indicando que la acción requiere justificación o modificación del destinatario.
5. El **Usuario** hace clic en "Override".
6. El sistema muestra un formulario de justificación.
7. El **Usuario** introduce la justificación.
8. El sistema registra el evento y permite el envío.

Alternativas de proceso y excepciones:

5a. El **Usuario** elimina al destinatario externo con “Remove recipient” y lo sustituye por uno interno:

- 5a1. El sistema detecta que ya no hay destinatarios externos.
- 5a2. El sistema permite el envío automáticamente.
- 5a3. El sistema vuelve al paso 8.

5b. El **Usuario** ignora el Policy Tip y envía el correo:

- 5b1. El sistema bloquea el envío del correo.

6.1.2.2.2 Insider Risk Management

A continuación, se detallan los casos de uso referentes a las políticas de evaluación de riesgos internos

6.1.2.2.2.1 Casos de uso

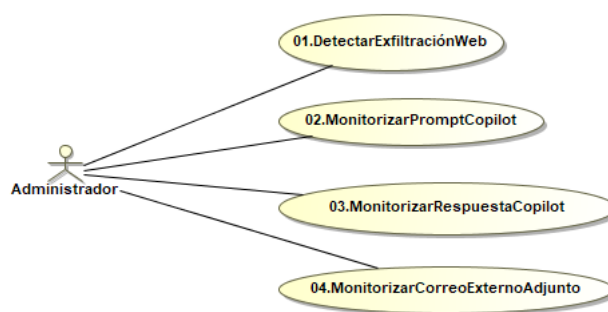


Figura 6. Diagrama de casos de uso Insider Risk Management

Caso de uso 01. DetectarExfiltraciónWeb

Resumen de la funcionalidad: Permite detectar si un usuario sube archivos a dominios web no autorizados desde un navegador, generando una alerta de riesgo si el contenido es sensible.

Parámetro de entrada: Actividad de un usuario desde un dispositivo registrado.

Parámetro de salida: Alerta generada en Insider Risk Management.

Actores: **Administrador**.

Precondición: El dispositivo desde el que se realiza la actividad debe estar registrado en Microsoft Purview.

Postcondición: Se genera una alerta si el sistema detecta que se ha subido información a un dominio no permitido.

Proceso normal principal:

1. Un usuario accede a un sitio web desde el navegador.
2. Un usuario sube uno o más archivos a un dominio externo.
3. El Sistema analiza el destino del tráfico web y detecta que el dominio no pertenece a la lista de confianza.
4. El Sistema analiza el contenido de los archivos y contiene información definida como sensible.

5. El Sistema registra una alerta si supera los umbrales establecidos y clasifica el nivel de riesgo según el tipo de información y la cantidad de eventos generados en el pasado.
6. El **Administrador** revisa la alerta y determina si se produjo o no exfiltración web.

Alternativas de proceso y excepciones:

4a. El archivo no contiene información sensible:

4a1. El sistema registra el evento.

4a2. El sistema vuelve al paso 5.

3a. El dominio pertenece a la lista de confianza:

3a1. El sistema no registra ningún evento de alerta.

Caso de uso 02. MonitorizarPromptCopilot

Resumen de la funcionalidad: Monitoriza si un usuario envía información sensible a través de Microsoft Copilot.

Parámetro de entrada: Prompt a Copilot.

Parámetro de salida: Alerta de riesgo.

Actores: **Administrador**.

Precondición: Copilot debe estar como indicador de una política en IRM.

Postcondición: Se genera alerta si se detecta compartición de contenido sensible.

Proceso normal principal:

1. Un usuario escribe un prompt a Copilot.
2. El sistema analiza el contenido del prompt.
3. Se detecta tipo de información sensible.
4. Se registra el evento en Insider Risk Management.
5. Se genera una alerta si supera el umbral establecido.
6. El **Administrador** revisa la alerta y determina si existe exfiltración real.

Alternativas y excepciones:

3a. Si no se detecta información sensible, no se genera evento.

Caso de uso 03. MonitorizarRespuestaCopilot

Resumen de la funcionalidad: Monitoriza si un usuario recibe información sensible a través de Microsoft Copilot.

Parámetro de entrada: Respuesta de Copilot.

Parámetro de salida: Alerta de riesgo.

Actores: **Administrador**.

Precondición: Copilot debe estar como indicador de una política en IRM.

Postcondición: Se genera alerta si se detecta respuesta con contenido sensible.

Proceso normal principal:

1. Un usuario escribe recibe un prompt de Copilot que contiene información sensible.
2. El sistema analiza el contenido de la respuesta.
3. Se detecta tipo de información sensible.

4. Se registra el evento en Insider Risk Management.
5. El **Administrador** revisa la alerta y determina si hubo o no exfiltración.

Alternativas y excepciones:

3a. Si no se detecta información sensible, no se genera ningún evento.

Caso de uso 04. MonitorizarCorreoExternoConAdjunto

Resumen de la funcionalidad: Permite monitorizar el envío de correos electrónicos con archivos adjuntos a destinatarios externos, supervisando si contienen información sensible para activar alertas según el nivel de riesgo.

Parámetro de entrada: Correo con archivos adjuntos.

Parámetro de salida: Correo enviado y evaluación del riesgo.

Actores: **Administrador**.

Precondición: El destinatario del correo no pertenece a la organización.

Postcondición: El correo es enviado y, si se cumplen ciertas condiciones, se genera una alerta en el entorno de Insider Risk Management.

Proceso normal principal:

1. Un usuario redacta un correo dirigido a un destinatario externo, adjunta uno o más archivos al correo y lo envía.
2. El sistema detecta que el correo contiene archivos adjuntos y que el destinatario es externo.
3. El sistema evalúa si el contenido adjunto está etiquetado como sensible o contiene información definida en los Sensitive Info Types.
4. El sistema detecta que contiene información sensible.
5. El sistema evalúa el dominio de destino.
6. El sistema detecta que el dominio de destino no está en la lista de confianza.
7. El sistema clasifica el nivel de riesgo (bajo, medio o alto) dependiendo de la cantidad de eventos realizados previamente y genera la alerta correspondiente.
8. La alerta se registra para revisión en las alertas de IRM.
9. El **Administrador** revisa la alerta con los eventos y determina si hubo o no riesgo.

Alternativas de proceso i excepciones:

4a. El adjunto no contiene información sensible:

4a1. El sistema vuelve al paso 5.

6a. El dominio pertenece a la lista de confianza

4b. El adjunto no contiene información sensible:

4b1. El sistema evalúa el dominio de destino.

4b2. El sistema detecta que el dominio de destino está en la lista de confianza.

6.1.2.2.3 Data Lifecycle Management

A continuación, se presentan los casos de uso para la solución implementada con esta solución de Microsoft para la gestión del ciclo de vida de los datos.

6.1.2.2.3.1 Casos de uso

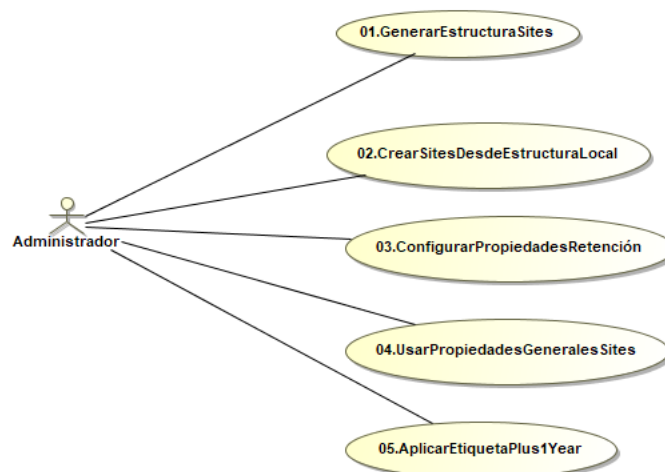


Figura 7. Diagrama de casos de uso Datalifecycle Management

Caso de uso 01. GenerarEstructuraSites

Resumen de la funcionalidad: Permite recorrer un directorio local con documentos PDF y clasificarlos automáticamente por país, tipo de trámite y cliente, generando una estructura de carpetas organizada para su posterior subida a SharePoint.

Parámetro de entrada: Directorio local con documentos PDF.

Parámetro de salida: Estructura de carpetas en local (Arbol_Sites).

Actores: **Administrador del sistema.**

Precondición: Deben existir documentos PDF almacenados en carpetas desorganizadas.

Postcondición: Se ha generado una estructura ordenada lista para ser convertida en Sites de SharePoint.

Proceso normal principal:

1. El **Administrador** ejecuta el script `extract_fields.py`.
2. El sistema analiza el texto de cada PDF buscando coincidencias con los campos definidos.
3. El sistema detecta el país, tipo de trámite y nombre del cliente.
4. El sistema crea una estructura de carpetas con la jerarquía País_Trámite > Cliente > Documentos.
5. El sistema copia los documentos a sus respectivas carpetas en Arbol_Sites.

Alternativas de proceso y excepciones:

2a. El PDF no contiene coincidencias válidas:

2a1. El sistema omite el archivo.

2a2. El sistema pasa al siguiente archivo y vuelve al paso 3.

Caso de uso 02. CrearSitesDesdeEstructuraLocal

Resumen de la funcionalidad: Permite crear Sites de SharePoint y sus bibliotecas (“Clients” y “Pending Appointments”) a partir de la estructura de carpetas generada previamente.

Parámetro de entrada: Carpeta con un árbol de directorios.

Parámetro de salida: Sites de SharePoint creados.

Actores: **Administrador del sistema.**

Precondición: La carpeta de árbol de directorios debe existir y contener subdirectorios organizados por país y cliente.

Postcondición: Los Sites y sus bibliotecas están disponibles en SharePoint con los documentos cargados.

Proceso normal principal:

1. El **Administrador** ejecuta el script create_sites.ps1.
2. El sistema recorre las carpetas de Arbol_Sites.
3. Para cada carpeta de país, crea un Site en SharePoint si no existe.
4. Dentro del Site, crea las bibliotecas “Clients” y “Pending Appointments” si no existen.
5. Sube los documentos de cada cliente a su carpeta correspondiente.

Alternativas de proceso y excepciones:

3a. El Site ya existe:

- 3a1. El sistema omite la creación y solo actualiza documentos.

5a. Error en la subida de archivo:

- 5a1. El sistema muestra mensaje de error pero continúa con el siguiente documento.

Caso de uso 03. ConfigurarPropiedadesRetención

Resumen de la funcionalidad: Permite asignar las propiedades mapeables GeneralRetention y SpecialRetentionLabel a los Sites creados en SharePoint para permitir su inclusión en scopes adaptativos en las políticas de retención de Data Lifecycle Management.

Parámetro de entrada: Archivo CSV con URL del Site y pares clave-valor de propiedades.

Parámetro de salida: Actualización de Sites con sus propiedades aplicadas.

Actores: **Administrador del sistema.**

Precondición: Los Sites deben estar creados y accesibles desde SharePoint.

Postcondición: Las propiedades mapeables han sido aplicadas a cada Site.

Proceso normal principal:

1. El Administrador ejecuta el script config_sites.ps1.
2. El sistema abre el CSV con la configuración.
3. Para cada línea, se conecta al Site indicado.
4. Aplica la propiedad clave:valor al Site, por ejemplo: GeneralRetention: 3years.
5. Registra el resultado de cada operación, si fue éxito o fallo, en el log.

Alternativas de proceso i excepciones:

3a. El Site indicado en el CSV no existe:

- 3a1. El sistema registra un error y continúa con el siguiente Site.

4a. La clave ya existe y overwrite está desactivado:

- 4a1. El sistema omite el cambio y registra el evento.

Caso de uso 04. UsarPropiedadesGeneralesSites

Resumen de la funcionalidad: El uso de las propiedades refinables creadas después de ser mapeadas a propiedades administrables permite definir scopes adaptativos para grupos de Sites en base a propiedades refinables y aplicarles políticas de retención automáticas.

Parámetro de entrada: Ninguno

Parámetro de salida: Scope creado y política de retención activa.

Actores: **Administrador del sistema.**

Precondición: Los Sites deben tener sus propiedades refinables asignadas correctamente.

Postcondición: Los Sites están bajo control de una política de retención específica.

Proceso normal principal:

1. El Administrador accede a Microsoft Purview.
2. Define un nuevo Adaptive Scope utilizando una query sobre el valor de RefinableString00.
3. El sistema valida que el Scope lista los Sites correctos.
4. El **Administrador** crea una política de retención para ese Scope.
5. La política se aplica a todos los Sites dentro del Scope automáticamente.

Alternativas de proceso y excepciones:

2a. La query es incorrecta

- 2a 1. El sistema no lista ningún Site

Caso de uso 05. AplicarEtiquetaPlus1Year

Resumen de la funcionalidad: Permite aplicar manualmente la etiqueta de retención “Plus1Year” a un documento en un Site de SharePoint para extender su tiempo de retención por un año adicional respecto a la política general aplicada al Site.

Parámetro de entrada: Documento en SharePoint.

Parámetro de salida: Etiqueta de retención “Plus1Year” aplicada.

Actores: **Usuario**

Precondición: El documento debe estar almacenado en un Site de SharePoint y debe tener asignada la propiedad RefinableString01=”1year”.

Postcondición: El documento pasa a tener una retención extendida de un año adicional.

Proceso normal principal:

1. El usuario accede a la biblioteca de documentos del Site desde SharePoint Online.
2. El usuario selecciona un documento que está próximo a cumplir su periodo de retención.
3. El usuario abre el menú de opciones del documento y selecciona “Aplicar etiqueta”.
4. El sistema muestra la lista de etiquetas disponibles.
5. El usuario selecciona la etiqueta “Plus1Year”.
6. El sistema aplica la etiqueta “Plus1Year” al documento seleccionado.
7. El sistema actualiza la propiedad de retención y registra el evento en Purview.

Alternativas de proceso y excepciones: Ninguna.

6.1.3 Fase aplicación web

6.1.3.1 Reglas del negocio

01. No puede reservarse una cita sin que exista previamente disponibilidad horaria asociada al trámite seleccionado.
02. Cada cita debe estar asignada a un único abogado disponible para la franja horaria seleccionada.
03. Todos los campos del formulario son obligatorios.
04. El país seleccionado debe coincidir con uno de los sitios configurados en SharePoint.
05. No se puede continuar con la reserva sin que el sistema registre correctamente el pago.
06. El token JWT debe generarse usando el nombre del cliente como base de autenticación tras el pago.
07. El JWT debe almacenarse en una cookie segura y HTTP-only accesible solo por el backend.
08. La cookie generada por el sistema debe estar disponible durante todas las operaciones posteriores de subida de documentos a SharePoint y agendar evento en calendario del abogado.
09. Si no se genera la cookie JWT, el sistema no permite agendar una cita ni generar documentos o subir archivos a SharePoint.
10. La generación de documentos debe ejecutarse antes de subir los archivos a SharePoint.
11. Los documentos deben incluir la información del cliente, del proyecto, del servicio contratado y del abogado asignado.
12. Los documentos generados deben subirse a la biblioteca 'Pending Appointments' dentro del Site de SharePoint correspondiente al país del trámite.
13. Si la subida a SharePoint falla, no debe continuarse con el agendado ni el envío de correos.
14. El sistema debe enviar un correo de confirmación al abogado y otro al cliente una vez creada la cita.

6.1.3.2 Especificación textual

En el siguiente apartado se detallan los casos de uso referentes al desarrollo de la aplicación web.

6.1.3.2.1 Casos de uso

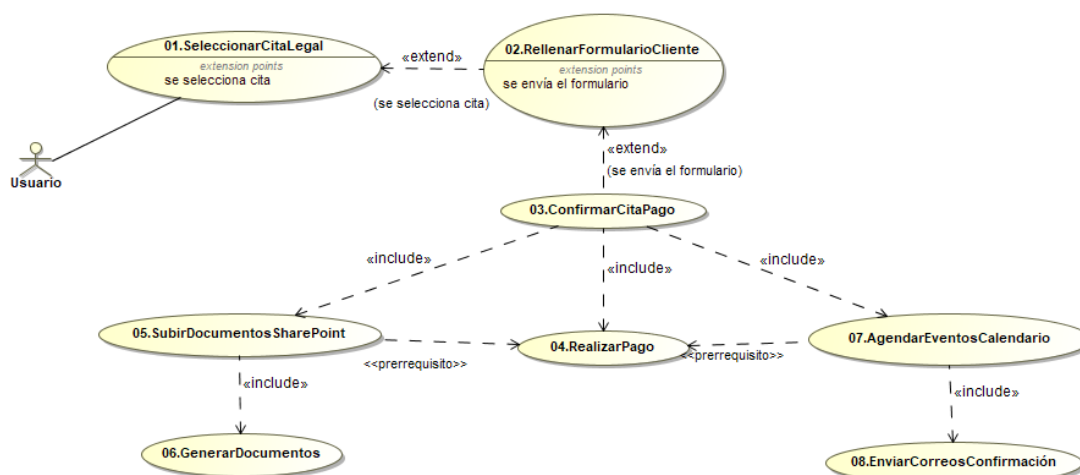


Figura 8. Diagrama de casos de uso de la aplicación web

Caso de uso 01. SeleccionarCitaLegal

Resumen de la funcionalidad: Permite a un cliente seleccionar una fecha y franja horaria disponible para una cita legal con un abogado.

Parámetro de entrada: Servicio legal deseado.

Parámetro de salida: Fecha y franja horaria seleccionada.

Actores: **Cliente.**

Precondición: Debe haber disponibilidad horaria en el calendario de los empleados del trámite deseado cargado desde el backend.

Postcondición: El sistema guarda temporalmente la selección del cliente.

Proceso normal principal:

1. El sistema muestra los días disponibles para el servicio legal seleccionado acompañado del abogado disponible.
2. El cliente selecciona una fecha.
3. El sistema muestra las franjas de horas disponibles para esa fecha.
4. El cliente selecciona una franja horaria y abogado disponible.
5. El sistema guarda la selección de forma temporal y avanza al siguiente paso.

Alternativas de proceso y excepciones:

1a. No hay disponibilidad para el servicio:

- 1a1. El sistema muestra un mensaje de “No hay fechas disponibles”.

Caso de uso 02. RellenarFormularioCliente

Resumen de la funcionalidad: Permite al cliente introducir sus datos personales y los de su proyecto para agendar la cita.

Parámetro de entrada: Datos personales y del proyecto.

Parámetro de salida: Campos del formulario completados.

Actores: **Cliente.**

Precondición: El **Cliente** debe haber seleccionado una fecha y franja horaria.

Postcondición: La información queda recolectada y lista para confirmación.

Proceso normal principal:

1. El sistema muestra el formulario con los campos a rellenar.
2. El cliente introduce: nombre, DNI, correo, dirección, país, nombre del proyecto y descripción de la idea.
3. El sistema valida que los campos estén completos.
4. El cliente hace clic en “Continuar” y el sistema guarda la información para confirmación.

Alternativas de proceso i excepciones:

3a. Algún campo está incompleto:

- 3a1. El sistema desactiva el botón “Continuar”.

Caso de uso 03. ConfirmarCitaPago

Resumen de la funcionalidad: Permite al cliente confirmar la cita y ejecutar todos los pasos asociados a la reserva.

Parámetro de entrada: Datos del cliente y la cita.

Parámetro de salida: Documentos generados, evento creado, correo enviado.

Actores: **Cliente**.

Precondición: Se deben haber completado correctamente los datos.

Postcondición: La cita queda registrada en todos los sistemas y los documentos clasificados.

Proceso normal principal:

1. El sistema muestra el resumen de cita y cliente.
2. El cliente confirma la cita.
3. El sistema ejecuta el caso de uso 04.RealizarPago
4. El sistema ejecuta el caso de uso 07.AgendarEventoCalendario
5. El sistema ejecuta el caso de uso 05.SubirDocumentosSharePoint
6. El sistema muestra un mensaje de que la cita se ha realizado con éxito

Alternativas de proceso y excepciones:

6a. El sistema muestra error y no agenda la cita.

Caso de uso 04.RealizarPago

Resumen de la funcionalidad: Simula el pago de la cita por parte del cliente generando un token JWT a partir del nombre del cliente.

Parámetro de entrada: Nombre del cliente.

Parámetro de salida: Confirmación de pago y token JWT en cookie.

Actores: **Sistema**.

Precondición: El cliente ha hecho clic en “Confirmar cita y pagar”.

Postcondición: El sistema marca la cita como pagada y genera la cookie de autenticación.

Proceso normal principal:

1. El sistema recibe el nombre del cliente tras la confirmación de la cita.

2. El sistema genera un token JWT firmado que representa la identidad del cliente autenticado.
3. El JWT se almacena en una cookie segura y HTTP-only, asociada al cliente.
4. El sistema devuelve la cookie al cliente.
5. El sistema marca la cita como pagada.

Alternativas de proceso y excepciones:

5a. Error interno:

- 3a1. El sistema informa que no ha sido posible procesar el pago.

Caso de uso 05. SubirDocumentosSharePoint

Resumen de la funcionalidad: Permite subir automáticamente los documentos de la cita al Site de SharePoint correspondiente.

Parámetro de entrada: Documentos y país del trámite.

Parámetro de salida: Documentos subidos al Site de SharePoint correspondiente.

Actores: **Sistema**.

Precondición: Los datos del formulario deben estar completados y debe haberse ejecutado correctamente el caso de uso 04.RealizarPago.

Postcondición: Los archivos se almacenan en el Site de SharePoint adecuado.

Proceso normal principal:

1. El sistema determina el Site según el país y el trámite.
2. Ejecuta el caso de uso 06.GenerarDocumentos
3. Conecta a SharePoint mediante la Graph API.
4. Sube el acuerdo y la factura generados en el paso 2. a la carpeta correspondiente.
5. Retorna status 200 ok

Alternativas y excepciones:

5a. Retorna error

Caso de uso 06. GenerarDocumentos

Resumen de la funcionalidad: Genera el acuerdo de confidencialidad y la factura del trámite realizado.

Parámetro de entrada: Datos del cliente y proyecto.

Parámetro de salida: Documentos.

Actores: **Sistema**.

Precondición: Se deben haber recibido los datos completos.

Postcondición: Los documentos se generan correctamente y están listos para ser enviados.

Proceso normal principal:

1. El sistema toma los datos enviados desde el frontend.

2. Genera un documento de acuerdo de confidencialidad con el proyecto a registrar entre el autor y el abogado con el que agenda la cita utilizando las plantillas del sistema y rellenando los campos target.
3. Genera un documento de la factura del servicio haciendo el mismo proceso que el paso anterior.
4. Los documentos son retornados con éxito.

Alternativas y excepciones: Ninguna

Caso de uso 07. AgendarEventoCalendario

Resumen de la funcionalidad: Permite registrar la cita en el calendario del abogado en Microsoft 365 y envía un correo de confirmación al abogado y al cliente.

Parámetro de entrada: Fecha, hora, abogado, cliente, correos.

Parámetro de salida: Evento creado.

Actores: **Sistema**.

Precondición: Debe haberse ejecutado el caso de uso 04.RealizarPago

Postcondición: El evento aparece en el calendario del abogado con los detalles de la cita y tanto el abogado como el cliente reciben correo de confirmación.

Proceso normal principal:

1. El sistema busca al empleado en el tenant.
2. Define hora de inicio y fin de la cita.
3. Crea un evento en su calendario.
4. Ejecuta el caso de uso 08.EnviarCorreosConfirmación
5. Retorna el éxito de la creación.

Alternativas y excepciones:

5a1. Retorna error de creación

Caso de uso 08. EnviarCorreosConfirmacion

Resumen de la funcionalidad: Envía automáticamente un correo de confirmación tanto al cliente como al abogado.

Parámetro de entrada: Email del cliente y abogado.

Parámetro de salida: Correos enviados.

Actores: **Sistema**.

Precondición: La cita ha sido agendada.

Postcondición: Ambos reciben los detalles por correo electrónico.

Proceso normal principal:

1. El sistema construye el contenido del mensaje para el abogado y el cliente.
2. Se envía el correo al abogado.
3. Se envía el correo al cliente.
4. Se registra el envío como completado.

Alternativas y excepciones:

4a1. Se retorna error.

7 Requisitos no funcionales

1. Toda la comunicación con Microsoft Graph debe realizarse a través de HTTPS.
2. El frontend debe desarrollarse en React utilizando TypeScript.
3. El backend debe estar implementado en Java con Spring Boot.
4. El backend debe exponer una API RESTful para el consumo desde el frontend.
5. Los scripts de automatización deben utilizar PowerShell y el módulo PnP excepto el de creación de directorios y extracción de campos que estará hecho con Python.
6. La infraestructura basada en Microsoft 365 debe permitir escalar sin necesidad de hardware adicional.
7. Toda la información etiquetada con las etiquetas Confidential e Industrial Property Use debe estar cifrada.
8. El sistema debe guiar al usuario a través de los pasos de la reserva mediante componentes visuales claros.
9. El sistema debe estar desarrollado en arquitectura desacoplada para facilitar su mantenimiento y evolución.
10. La integración con Microsoft Graph debe realizarse mediante Microsoft Graph SDK y autenticación con client credentials.

8 Tecnologías empleadas en protección de datos: Microsoft Purview

Microsoft Purview es la plataforma unificada de gobernanza, protección de la información y cumplimiento normativo de Microsoft 365. Su objetivo es permitir a las organizaciones proteger los datos sensibles, gestionar su ciclo de vida, garantizar el cumplimiento normativo y supervisar actividades de riesgo.

En este Trabajo de Fin de Grado se han utilizado diversas capacidades de Microsoft Purview. A continuación, se describen de forma técnica los componentes clave empleados en la solución:

8.1 Etiquetas de sensibilidad (Sensitivity Labels)

Las etiquetas de sensibilidad permiten a las organizaciones clasificar y proteger la información en función de su nivel de confidencialidad. Una etiqueta puede aplicarse manualmente por el usuario, sugerirse de forma recomendada o asignarse automáticamente en base al contenido.

Las etiquetas permiten:

- Aplicar cifrado a correos y documentos.
- Restringir accesos por usuario, grupo o dominio.
- Insertar marcas visuales (cabeceras, pies de página, marcas de agua).
- Activar o bloquear funciones como reenviar, copiar/pegar o imprimir.

Las etiquetas se gestionan desde Microsoft Purview y se aplican en entornos como Outlook, Word, Excel, SharePoint, OneDrive o Teams.

8.2 Tipos de información sensible (Sensitive Info Types)

Un tipo de información sensible (SIT) es un conjunto de patrones, expresiones regulares, palabras clave o lógica avanzada que permite identificar automáticamente información de alto valor o confidencial.

Microsoft proporciona cientos de tipos predefinidos (como números de tarjeta, DNI, IBAN, pasaportes...), y también permite crear SIT personalizados, combinando:

- Palabras clave.
- Patrones regex.
- Confianza (confidence level).

Los SIT son la base de detección en etiquetas automáticas, políticas DLP y recomendaciones de clasificación, definiendo así la información sensible de una organización.

8.3 Prevención de pérdida de datos (Data Loss Prevention - DLP)

La prevención de pérdida de datos (DLP) permite detectar y controlar actividades en las que se expone información sensible, como:

- Envío de correos a dominios externos.
- Carga de archivos en ubicaciones no autorizadas.
- Uso indebido de contenido con datos protegidos.

Una política DLP puede configurarse para:

- Notificar al usuario su comportamiento inadecuado mediante un cuadro de aviso “Policy Tip”.
- Solicitar justificación para continuar con una acción que puede poner en riesgo la compañía.
- Bloquear completamente un determinado tipo de acción.

Una política DLP permite definir condiciones basadas en tipos de información sensible, etiquetas de sensibilidad o metadatos. También es posible personalizar reglas por ubicación (Exchange, SharePoint, Teams, etc.) En este trabajo de fin de grado nos focalizamos en Exchange.

En conclusión, lo más importante es diferenciar que las políticas DLP tienen como finalidad de detectar comportamientos inadecuados y configurar acciones que mitiguen estos comportamientos.

8.4 Data Lifecycle Management (DLM)

La gestión del ciclo de vida de los datos permite controlar la conservación y eliminación de documentos en función de criterios organizativos, legales o normativos. Las políticas de retención pueden aplicarse:

- Por tiempo desde la creación o modificación.
- Por aplicación de una etiqueta de retención.
- Por inclusión del contenido en un ámbito adaptativo (Adaptive Scope).

Cuando un documento cumple su período de retención, puede:

- Ser eliminado automáticamente.
- Ser revisado.
- Ser retenido por motivos legales.

Este sistema asegura que la información no se almacena indefinidamente y permite aplicar estrategias de limpieza y cumplimiento sin intervención manual.

8.5 Adaptive Scopes

Los Adaptive Scopes permiten aplicar políticas de forma dinámica a un conjunto de recursos (usuarios, grupos, sitios de SharePoint) basándose en atributos o propiedades.

En lugar de seleccionar manualmente los objetos a los que se aplica una política, se define una consulta lógica sobre sus propiedades. Por ejemplo:

Aplicar una política solo a los Sites de SharePoint donde RefinableString00 = OneYear.

Esto proporciona una gestión escalable y flexible. Los scopes se actualizan automáticamente cuando nuevos objetos coinciden con la condición.

8.6 Refinable Strings (Propiedades gestionables en SharePoint)

Las Refinable Strings son propiedades gestionadas en SharePoint Online que permiten clasificar, buscar, filtrar y aplicar políticas sobre el contenido almacenado. Son especialmente útiles cuando se necesita vincular recursos como documentos o sitios a políticas de retención o seguridad basadas en metadatos.

Técnicamente, una Refinable String es una Managed Property de tipo string predefinida por Microsoft, identificada como 'RefinableString00', 'RefinableString01', hasta 'RefinableString199'. Estas propiedades pueden ser utilizadas en:

- Scopes adaptativos (Adaptive Scopes) para aplicar dinámicamente políticas de retención (Es el caso de uso de este TFG).
- Microsoft Search para habilitar filtros refinados en búsquedas.
- Consulta de contenido en Purview, para clasificación y gobernanza automatizada.

8.6.1 Crawled Properties

Para poder utilizar una Refinable String, es necesario comprender el concepto de Crawled Properties. Estas son propiedades detectadas automáticamente por el sistema durante el rastreo del contenido en SharePoint y deben ser vinculadas a una Refinable String para poder ser utilizadas. Pueden ser nombres de propiedades de archivos o campos creados por el usuario.

El proceso de habilitación de una Refinable String para aplicar en las políticas de retención de DLM consiste en:

1. Crear una crawled property para el propósito deseada. Estas crawled properties serán del estilo clave:valor, por ejemplo: 'GeneralRetention:OneYear'
2. Mapear esa propiedad a una RefinableStringXX desde el Centro de administración de SharePoint (Search Schema).
3. A partir de ese momento, la propiedad refinada puede utilizarse en filtros de búsqueda, políticas de retención o lógica en Microsoft Purview.

Este mapeo convierte una propiedad no estructurada (crawled) en una propiedad controlada y utilizable de forma estructurada (managed) y se habilita así para poder ser utilizada en Scopes Adaptativos e incluir dichos Scopes en políticas de retención.

8.7 Insider Risk Management (IRM)

Insider Risk Management es una solución incluida en Microsoft Purview que permite detectar y evaluar actividades internas potencialmente riesgosas realizadas por empleados o usuarios dentro de la organización. Puede pensarse que es parecido a soluciones como DLP, pero IRM no bloquea acciones de forma directa, sino que proporciona visibilidad y contexto sobre comportamientos que pueden suponer un riesgo para la seguridad de los datos.

El modelo de IRM se basa en un sistema de evaluación progresiva del riesgo, estructurado en varias capas que se complementan, estas son las siguientes:

8.7.1 Eventos desencadenantes (Triggering Events)

Los eventos desencadenantes son acciones iniciales que hacen que el sistema comience a monitorizar de forma intensiva la actividad de un usuario específico. Estos eventos no generan una alerta por sí mismos, pero indican que hay un posible contexto de riesgo.

Algunos ejemplos comunes de triggering events son:

- Acceso masivo a archivos confidenciales.
- Envío de correos con archivos adjuntos a dominios externos.

A partir del evento desencadenante, el sistema inicia un periodo de observación donde se analizan en detalle las actividades del usuario para contemplar si se debe generar una alerta o forma parte de una actividad legítima para la empresa. Estos límites debe definirlos la empresa.

8.7.2 Indicadores de riesgo

Durante el periodo de observación, IRM recopila indicadores de riesgo, que son métricas acumulativas que miden la repetición en las acciones de un usuario y el contexto de ciertas acciones sensibles. Estos indicadores están definidos por Microsoft y el empresario debe seleccionar cuál se adapta mejor para su caso de estudio. Estos indicadores incluyen actividades como:

- Carga de archivos en aplicaciones no autorizadas (Google Drive, WhatsApp Web, etc.).
- Transferencias masivas de datos fuera de horario laboral.
- Manipulación de documentos etiquetados como confidenciales.
- Acciones sospechosas en dispositivos no corporativos.

Cada indicador tiene un umbral de activación definible. Cuando un usuario supera uno o varios de estos umbrales, se considera que el comportamiento ha escalado en riesgo.

8.7.3 Clasificación de alertas y detección inteligente

Cuando se superan los umbrales definidos, el sistema genera una alerta, la cual es clasificada en niveles de bajo, medio o alto riesgo, según los siguientes factores:

- Número total de indicadores activados.
- Gravedad individual de los indicadores.
- Frecuencia y distribución temporal.
- Contexto (por ejemplo, si el usuario está en periodo de salida de la empresa).

Adicionalmente, Microsoft Purview incluye una capa llamada Intelligent Detections, que permite priorizar automáticamente ciertas alertas en función de patrones complejos de comportamiento, incluso si no se han superado todos los umbrales numéricos. Estas detecciones inteligentes se basan en análisis heurísticos y aprendizaje automático sobre la actividad organizativa global y la actividad de cada usuario de riesgo en concreto.

9 Diseño

En este apartado detallaré todas las estrategias seguidas y el diseño resultante de ellas que han permitido la arquitectura de la solución propuesta.

9.1 Diseño fase usuarios híbridos

Como se planteó en la introducción, para poder habilitar identidad híbrida se necesita un Active Directory y un tenant de Microsoft Azure.

El Active Directory es posible gracias a la utilización de un controlador de dominio y máquinas unidas a ese dominio. Es por este motivo que en este diseño se he optado por construir un entorno compuesto por máquinas virtuales que simulan una infraestructura empresarial básica. La estructura es la siguiente:

- Máquina virtual con Windows Server 2022 (WServer): actúa como controlador de dominio y aloja el servicio de Active Directory Domain Services (AD DS).
- Tres máquinas virtuales con Windows 11 Pro (WClienteN): funcionan como clientes unidos al dominio del controlador de dominio. Estas máquinas permiten simular distintos empleados accediendo a recursos corporativos.

El nombre de dominio configurado en el entorno es cdserver.local. Todas las máquinas virtuales han sido desplegadas mediante la tecnología Hyper-V.

9.2 Diseño de usuarios y grupos de M365

Para poder desarrollar una solución integral, he decidido diseñar una estructura básica de usuarios y grupos para un despacho de abogados de la propiedad intelectual.

9.2.1 *Diseño departamento de usuarios y licencias*

He decidido que cada usuario pertenece a un departamento en concreto de la empresa. Los departamentos son los siguientes:

- Departamento de Marcas
- Departamento de Patentes
- Departamento de Diseño Industrial
- Departamento de Modelo de Utilidad
- Departamento de Finanzas
- Departamento de Informática

Cada uno de estos usuarios tiene licencias de M365 para poder trabajar con el paquete de Office completo.

9.2.2 *Diseño de grupos*

Para poder desarrollar la solución con éxito he decidido crear un grupo de usuarios para poder simular un entorno de trabajo real. El grupo es el siguiente:

- Intellectual Property Agent: En este grupo se incluye todos los usuarios cuyo departamento esté relacionado con algún registro de la propiedad intelectual.

9.3 Diseño fase protección de datos

9.3.1 *Microsoft Sensitive Info Types*

Una organización debe definir qué datos son sensibles para ella. En base a la definición de esta información, podemos aplicar políticas y herramientas para detectarla y protegerla. En el caso de uso, son abogados de la propiedad intelectual, por ello se definieron tres tipos:

9.3.1.1 **Product_Registration:**

Esta información sensible se ha diseñado para que defina todos los títulos posibles para un registro en el ámbito de desarrollo del proyecto: Abogados de la Propiedad Intelectual.

Es decir, si un documento es generado en el OneDrive o SharePoint o está en formato .docx y contiene alguno de los títulos referentes a documentos de registro en Oficinas de Patentes y Marcas en la Unión Europea, España y Estados Unidos, sería posible detectarlo gracias a la utilización de esta definición de información sensible.

Para pertenecer a este tipo de información sensible, el documento debe contener alguno de los siguientes títulos:

- Marca nacional
- Solicitud de modelo de utilidad
- Solicitud de patente
- Mark Information
- RCD file information
- EUTM file information

Esto se ha configurado en Purview mediante una lista de palabras clave.

9.3.1.2 **Invention_Proposal**

Esta información sensible hace referencia a contratos de confidencialidad y propuestas de invenciones. Ambos casos los he diseñado en base a lo que he considerado correcto:

- **Contratos de confidencialidad:** Tienen de título “Contrato de confidencialidad” o “Confidential Agreement” y en su contenido contendrá los campos “Título”, “Autor” y “Descripción de la idea confidencial”, y sus respectivas versiones en inglés.
- **Propuestas de Invención:** Tienen de título “Propuesta de invención” o “Proposal for invention” y en su contenido contendrá los campos “Autor, Título”, “Descripción”, “Anexos” y sus respectivas versiones en inglés.

La configuración se ha llevado a cabo en Purview definiendo dos listas de palabras, cada una para los puntos mencionados anteriormente.

9.3.1.3 **Invoice**

Esta información sensible se ha diseñado de forma que haga referencia a documentos con contenido referente a facturas o a transacciones económicas que contengan divisas, esto debido a lo que puede suponer que se exfiltre un posible presupuesto o las tarifas que maneja la empresa.

De esta forma, se han diseñado tres patrones para que se detecte esta información:

- **pre_divisa_regex:** con este patrón detectamos patrones de divisas en un documento, figurando antes de la cifra, el símbolo de la moneda. Por ejemplo, si en un documento aparece el patrón “€50”, “\$50” el patrón lo detectaría.
- **final_divisa_regex:** con este patrón nuestra intención es de nuevo detectar información de montos de dinero en el documento, con la diferencia de que ahora el símbolo de la moneda puede estar detrás. Por ejemplo, si en un documento aparece el patrón “50€”, “50\$”, “50 euros” o “50 EUROS” el patrón lo detectaría.
- **invoice_list:** este patrón contiene una lista de términos comúnmente utilizados, como “Invoice, Factura, Rechnung, Facture” para el desarrollo de una transacción financiera.

9.3.2 Microsoft Information Protection

En la introducción de este trabajo de fin de grado, se comentó la importancia de tener la información clasificada como mecanismo de protección de la información y existen varias estrategias para realizarlo, en nuestro caso de uso hemos aplicado tres tipos de aplicación:

- Manual: El usuario aplica manualmente la etiqueta deseada.
- Recomendada: El usuario recibe un pop-up con una recomendación de aplicación de determinada etiqueta, dependiendo del contenido de la información que está manejando.
- Automática: En base a reconocimiento de patrones en un documento, el sistema es capaz de identificarlo y aplicar automáticamente una etiqueta con un tipo de cifrado y nivel de confidencialidad.

Para aplicar esas tres estrategias de forma integral, he diseñado cuatro etiquetas:

9.3.2.1 Confidential non encrypted:

Esta etiqueta se ha creado con la finalidad de que la información enviada no esté cifrada, por lo que cualquier persona podría tener acceso a ella. Su implementación está fuertemente vinculada a la política de prevención de pérdida de datos “DLP Unencrypted Label” que detecta el uso de la etiqueta y exige justificar el envío de información sin cifrar si es a un externo. Dicha política de prevención de pérdida de datos será explicada en el punto de diseño de Data Loss Prevention.

Se ha llevado a cabo un diseño de la etiqueta que requiera de la aplicación de forma manual ya que estoy siguiendo la estrategia de obligar al usuario a comprender la gravedad de sus acciones y la sensibilidad de la información que comparte externamente. Si un usuario desea enviar información sin cifrar a un destinatario externo a la organización, para que el destinatario pueda leer el mensaje y no esté bloqueado su contenido, deberá usar esta etiqueta, ya que las demás etiquetas disponibles cifran automáticamente su contenido cuando el destinatario es externo.

9.3.2.2 Confidencial Encrypted

Esta etiqueta se ha configurado para realizar una recomendación al usuario de aplicar esta etiqueta a documentos y correos si detecta que estos contienen información relativa a:

- IBAN
- DNI
- Número de pasaporte
- Tarjetas de débito/crédito
- Sensitive Info Type “Invoice”

En este caso, la estrategia que se ha seguido es de recomendación, es decir, si un usuario está redactando un correo o un documento y la herramienta detecta que se está usando la información anteriormente especificada, mediante un pop-up recomendará al usuario hacer un cambio de etiqueta. Se siguió esta estrategia ya que de esta manera se está involucrado directamente al trabajador en el uso de la etiqueta y en el entendimiento de la información que maneja.

Se espera que tras varios usos el usuario entienda que la información de este tipo personal que administra debe estar encriptada ya que es sensible y su pérdida puede suponer un costo muy grande a la empresa.

9.3.2.3 **Industrial Property Use**

He diseñado esta etiqueta de modo que si es aplicada a un documento o correo sólo tengan acceso a la información etiquetada los usuarios que pertenezcan al grupo de agentes “Intellectual Property Agent”.

La estrategia seguida en la configuración de esta etiqueta es que la etiqueta se aplicará automáticamente al detectar que se está redactando un correo o documento que contiene información sensible del tipo configurado en los tipos de información sensibles “Product_Registration” y “Invention_Proposal”.

Por lo tanto, si una persona está redactando un correo o un documento y comienza a redactar patrones que coincidan con la configuración definida, el sistema automáticamente aplicará la etiqueta Industrial Property Use.

En este caso se decidió que fuera una aplicación automática ya que en el desarrollo del presente trabajo de fin de grado se verá que hay una aplicación web interactuando con la generación de gran parte de estos documentos de forma automática, lo que ocasiona que muchas veces los empleados no interactúen inmediatamente con los documentos y aunque no lo hagan queremos proteger la información relacionada a ideas sin registrar y contratos de confidencialidad, por ello de forma automática evitamos errores y protegemos la información.

9.3.2.4 **Internal Use**

Esta etiqueta se ha diseñado con una función puramente informativa. Es la etiqueta aplicada por defecto en todos los documentos y correos en la empresa e indica que su uso es para temas internos a modo de clasificación.

Esta etiqueta también cifra la información compartida, pero se ha decidido que sería buena idea crearla ya que puede ayudar a un usuario a diferenciar que, si tiene esta etiqueta aplicada a su documento o correo, quiere decir que no está en ese momento manejando información sensible y por lo tanto permitir al usuario entender mejor sus acciones y sus consecuencias.

9.3.3 **Data Loss Prevention**

A continuación se explica la política de Prevención de Pérdida de datos implementada.

9.3.3.1 **DLP Unencrypted Label**

Como se mencionó en el apartado anterior, esta política de prevención de pérdida de datos se ha hecho con la estrategia de detectar si un correo está siendo enviado a un externo sin ningún tipo de cifrado, es decir, aplicando la etiqueta “Unencrypted Label”.

Gracias a esta política podemos detectar el caso de uso y hacer una acción al respecto, en este caso, la acción es solicitar una justificación para invalidar el bloqueo del correo y proporcionarnos esta estrategia dinámica que permite a los trabajadores entender las consecuencias de sus acciones.

9.3.4 Insider Risk Management

Se han diseñado dos políticas para detectar riesgos internos con dos estrategias diferenciadas:

9.3.4.1 File Uploads

Como su nombre indica, esta política está diseñada para detectar si los usuarios realizan exfiltraciones de documentos que contienen o no información sensible. Dado que el ámbito de estudio de este trabajo es un desarrollo lo más enfocado a la nube posible, he considerado únicamente exfiltraciones realizadas a través de aplicaciones web.

Esta política sólo detectará actividad proveniente de dispositivos asociados a cuentas de usuarios, por este motivo he incluido el grupo de usuarios “Intellectual Property Agent” ya que cada uno de estos usuarios tiene asociado una máquina virtual la cual está dada de alta en el tenant.

En vista del creciente uso de la inteligencia artificial en tareas de horario laboral, también he decidido hacer uso de los indicadores referentes a la detección de prompts con información sensible enviada y recibida por Copilot, ya que es común que los empleados desconozcan el riesgo que conlleva enviar datos confidenciales a una IA, por ello si se detecta debe ser evaluado y corregido.

9.3.4.2 External emails with attachments

Esta política se encarga de monitorizar los correos que se envíen a dominios externos con archivos adjuntos.

La finalidad de esta política es detectar cuándo una persona está enviando información al exterior, estando o no etiquetada y añadir un nivel de severidad dependiendo de su dominio destino.

9.3.5 Data Lifecycle Management con Adaptive Scopes

En el contexto de una firma de abogados especializada en propiedad intelectual, donde se gestionan grandes volúmenes de documentación legal con diferentes tiempos de vigencia, me resultó muy relevante diseñar una estrategia que permita la retención selectiva según el tiempo de renovación del documento, para poder tener así una limpieza automatizada de los ficheros que no se han modificado en mucho tiempo y asegurarnos de retener los ficheros por la cantidad de tiempo adecuada.

El diseño de las políticas de retención configuradas están enfocadas a Sites de SharePoint. A continuación, detallo el diseño seguido tanto de los Sites de SharePoint como de las políticas de retención con ámbito adaptativo.

9.3.5.1 Diseño de Sites de SharePoint

Como expliqué en la introducción de este trabajo, en parte del contexto explicado se plantea que el caso de uso involucra documentos almacenados de forma local de forma ineficiente y desordenada y se plantea la solución de migrar estos documentos a Sites de SharePoint de forma automatizada.

Para dar solución a esta automatización se diseñaron dos scripts:

- `extract_fields.py`: extrae información del directorio de documentos desordenado y crea la siguiente estructura de directorios:

```

Árbol_Sites
  + País_Trámite
    + Cliente
      - Registros del cliente
  
```

- `create_sites.ps1`: Recorre el árbol de directorios generado en el paso anterior y replica esa estructura en SharePoint:
 - 'País_Trámite' es el nombre del Site de SharePoint.
 - 'Cliente' es la carpeta del cliente que se creará en la biblioteca "Clients" del Site de SharePoint.
 - 'Registros del cliente' Será un directorio por cada registro dentro del cliente y dentro de este directorio irá el registro correspondiente.

Un ejemplo del resultado esperado seguiría esta estructura:

```

Sitio SharePoint
  +Clients
    + Nombre del cliente
      + Registro1
        -Fichero1 Registro1
        -Fichero2 Registro1
      
```

9.3.5.2 Diseño de Políticas de Retención con Adaptive Scope

Se diseñó un script para cumplir con el proceso de asignación de propiedades de forma automatizada para obtener la retención selectiva a cada sitio anteriormente mencionada. El script es el siguiente:

- `config_sites.ps1`: está diseñado para recibir las url de los Sites a asignar propiedades y la propiedad a asignar. En nuestro caso de uso se aplica la propiedad mapeable "GeneralRetention" bajo los siguientes Sites y periodos de retención:
 - Spain_Patents: tres años
 - Spain_Trademarks: doce años
 - Spain_UtilityModels: tres años
 - EuropeanUnion_IndustrialDesigns: tres años
 - EuropeanUnion_IndustrialDesigns: tres años
 - UnitedStates_Trademarks: doce años

La estrategia seguida para estos periodos se basa en la frecuencia de renovación de cada trámite, y a cada una de esas frecuencias, le he sumado dos años extra para dar un margen de tiempo de inactividad.

Asimismo, una patente en España tiene una frecuencia de renovación anual, por lo que he decidido que se dé un margen de inactividad de dos años adicionales a la fecha de renovación, para decidir finalmente que si en tres años no se ha hecho ningún cambio en ese

directorio de cliente es porque no sigue trabajando con la empresa o no va a renovar más su registro y por lo tanto llega a su fin su periodo de retención y es momento de borrarlo.

También he definido otra propiedad mapeable con la finalidad de ser utilizada en políticas de retención en etiquetas. Esta propiedad se ha denominado “SpecialRetentionLabel” y se usará para dar un año adicional a la duración general de la política de retención.

Estas propiedades mapeables serán asignadas a propiedades gestionables para poder ser utilizadas en el Adaptive Scope. He decidido mapear GeneralRetention con RefinableString00 y SpecialRetentionLabel con el RefinableString01.

Esas propiedades serán tratadas en un Adaptive Scope, publicadas en políticas de retención las cuales contendrán el scope en concreto.

El primer Adaptive Scope tendrá esta query de Microsoft:

- RefinableString00:ValorGeneralRetention

El segundo Adaptive Scope tendrá esta query de Microsoft:

- RefinableString01:1year

9.4 Diseño aplicación web

La siguiente sección detallará el diseño de la aplicación web. Primeramente se desarrollará el backend, y luego el frontend.

9.4.1 Backend

Para el desarrollo del backend he utilizado el framework Spring Boot, el cual es una solución moderna basada en el lenguaje Java que facilita la creación de aplicaciones empresariales con arquitectura REST. Este framework sigue el paradigma de modelo-vista-controlador (MVC) y promueve el uso de patrones de diseño estandarizados, lo que me permitió mantener un código modular, escalable y mantenible.

El resultado de este diseño es una API RESTful en Java cuyo objetivo principal es integrarse con la API de Microsoft Graph, permitiendo así realizar operaciones como:

- Consultar la disponibilidad horaria de los empleados.
- Obtener información de sus departamentos y horarios.
- Agendar eventos en sus calendarios de Microsoft 365.
- Subir documentos a sitios de SharePoint.
- Enviar correos confirmación.

No he necesitado incluir una base de datos estructurada en esta solución, ya que todos los datos requeridos son obtenidos directamente del formulario del cliente en el frontend o recuperados dinámicamente mediante las consultas a Microsoft Graph anteriormente mencionadas.

La explicación de la funcionalidad de cada parte de la API será explicada en el apartado de implementación de esta documentación.

9.4.2 Frontend

En el caso del frontend, he utilizado el framework React junto con el lenguaje TypeScript, ya que esto me pareció lo ideal para mi solución al tener tipado estático, mayor seguridad en tiempo de desarrollo y herramientas modernas para mantener una base de

código sólida. Hacer esta elección me permitió construir una interfaz gráfica limpia y minimalista, además de utilizar buenas prácticas y crear un código extensible, reutilizable y escalable.

El diseño del flujo de interacción del usuario lo he basado en un patrón multistep, compuesto por tres pasos principales: selección de cita, introducción de datos del cliente y confirmación. Cada uno de estos pasos lo he diseñado como un componente funcional, para permitir la reutilización de código.

Una parte crucial en el desarrollo de esta solución fue realizar la gestión del estado local de cada paso, utilizando UseState. Esto lo necesitaba porque me permite acumular progresivamente los datos introducidos por el usuario y mantener el estado a lo largo del flujo completo hasta el final, ya que una vez completado el proceso, los datos recolectados los utilizo para realizar las distintas llamadas a la API REST, como la generación de documentos, la subida a SharePoint o la creación de eventos en el calendario.

El diseño de la interfaz gráfica he decidido realizarlo con un estilo minimalista y práctico, ideal para mi proyecto ya que mi finalidad era desarrollar una solución integral que fuera fácil de utilizar y que cumpliera con el objetivo de forma limpia y eficiente.

A continuación, muestro capturas y explicación sobre el diseño de las vistas de la interfaz gráfica:

- **Vista de inicio:** en esta vista podemos observar todos los trámites que maneja la empresa de abogados de la propiedad intelectual.



Figura 9. Vista de inicio

Ahora se muestra la misma pantalla de inicio, con la diferencia de que el cliente ha puesto el cursor sobre el trámite que desea y ha aparecido el botón “Reservar cita”:



Figura 10. Vista de inicio con el cursor sobre el trámite deseado

- **Vista de inicio:** en esta vista mostramos las fechas disponibles de los abogados. Se muestra un mes desde el día actual en el que se está realizando la cita.

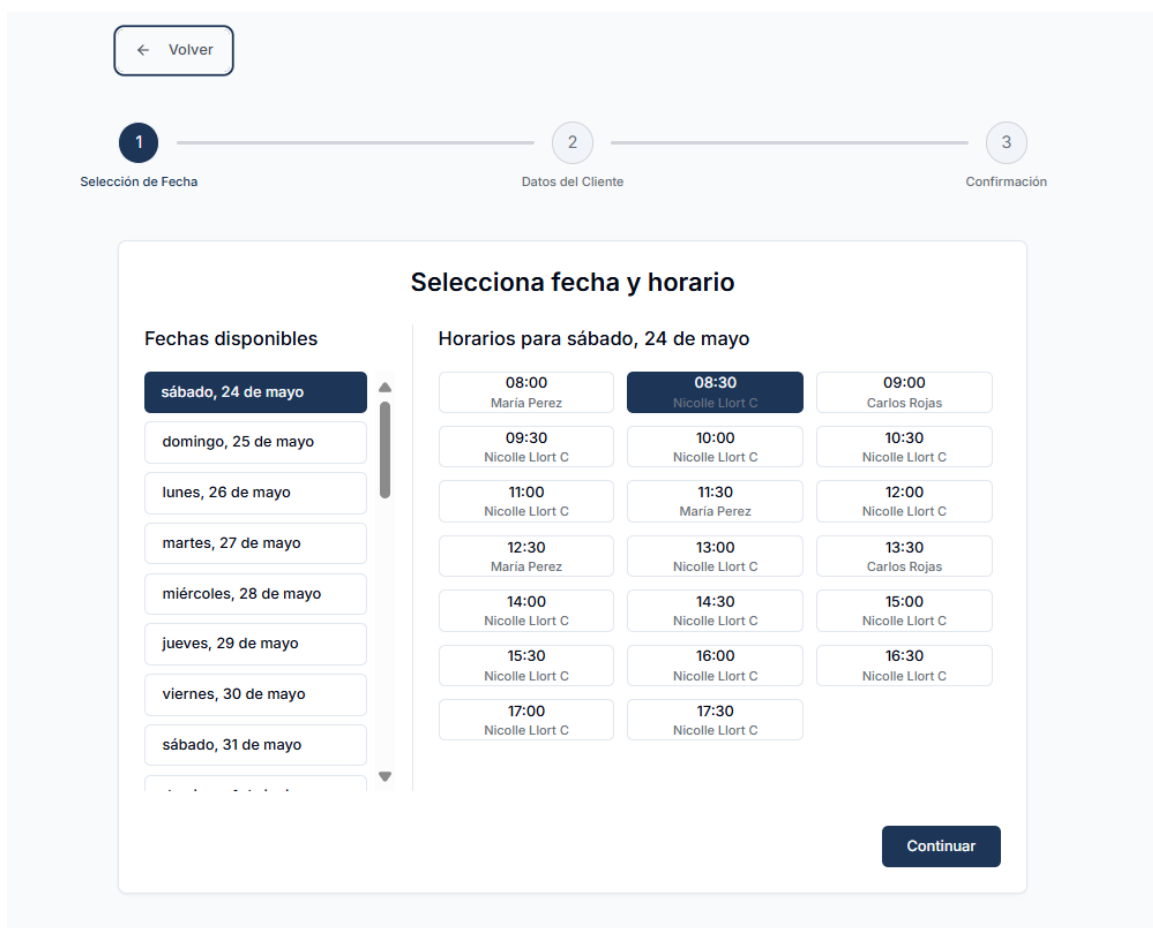


Figura 11. Vista de selección de franja horaria

- **Vista de formulario:** En esta vista se rellenan todos los datos personales del cliente y todo lo referente al proyecto que desea registrar. Todos los campos son

obligatorios, si no se rellenan todos, el campo de continuar no aparece en azul oscuro, como se puede ver en la Figura 12. En la Figura 14 podemos apreciar el selector de país de registro desplegado, para poder apreciar los detalles del mismo.

← Volver

1 Selección de Fecha 2 Datos del Cliente 3 Confirmación

Datos del Cliente y Proyecto

Nombre del autor: Paola Rodríguez

DNI del autor: 45663297R

Correo electrónico: paolarodriguez@gmail.com

País de registro: España

Dirección del autor: Calle Llovera, 15. 43205, Reus

Nombre del proyecto: Registro de marca - Paola TIC

Descripción de la idea: Quiero registrar mi marca como "Paola TIC" porque es un nombre que he visto que no está registrado en ningún sitio aún y quisiera quedarme yo con los derechos de esta marca ya que trabajo con productos del sector TIC y me gusta el nombre.

Volver Continuar

Figura 12. Vista de formulario

← Volver

1 Selección de Fecha 2 Datos del Cliente 3 Confirmación

Datos del Cliente y Proyecto

Nombre del autor: [Empty]

DNI del autor: [Empty]

Correo electrónico: [Empty]

País de registro: Selecciona un país

Dirección del autor: [Empty]

Nombre del proyecto: [Empty]

Descripción de la idea: [Empty]

Volver Continuar

Figura 13. Vista de formulario sin rellenar

Pais de registro

Selecciona un país ▾

- España
- Unión Europea
- Estados Unidos

Figura 14. Selector de país

- Vista de confirmación de cita:** en esta cita se muestran todos los detalles previamente rellenos. Es en este paso donde tras presionar el botón de “Confirmar cita y pagar” estamos realizando el pago ficticio, generando el token JWT y realizando todas las llamadas a la API correspondientes para reservar la cita con éxito.

✓
✓
3

Selección de Fecha
Datos del Cliente
Confirmación

Confirma tu Cita

Datos de la Cita

Servicio: Marcas	Abogado: Nicolle Llort C
Fecha: 24 de mayo de 2025	Hora: 08:30

Datos del Cliente

Nombre: Paola Rodríguez	DNI: 45663297R
Dirección: Calle Llovera, 15. 43205, Reus	

Datos del Proyecto

Nombre del Proyecto:
Registro de marca - Paola TIC

Tipo de Documento:

Descripción:
Quiero registrar mi marca como "Paola TIC" porque es un nombre que he visto que no está registrado en ningún sitio aún y quisiera quedarme yo con los derechos de esta marca ya que trabajo con productos del sector TIC y me gusta el nombre.

Volver
Confirmar Cita y Pagar

Figura 15. Vista de confirmación de cita

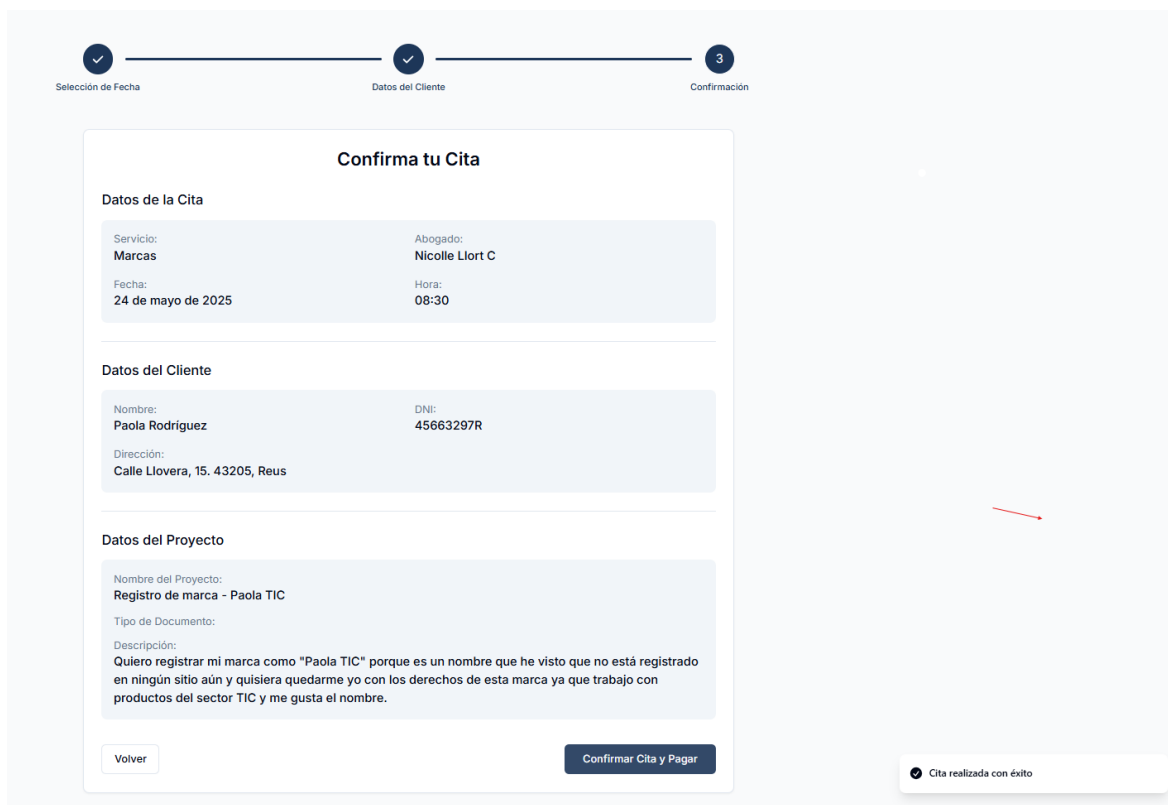


Figura 16. Vista de cita realizada con éxito

10 Implementación

10.1 Fase de usuarios híbridos

A continuación, detallaré la solución híbrida desarrollada en el presente Trabajo de Fin de Grado. En primer lugar, se detalla el proceso de creación de una red local que simula la de una organización, y posteriormente describiré la configuración necesaria para lograr la sincronización de usuarios entre el entorno local y la nube de Microsoft.

10.1.1 Creación de una red local

Para poder diseñar una solución basada en usuarios híbridos, es necesario simular un entorno de red con un directorio local de usuarios, utilizando Active Directory Domain Services (AD DS). Para ello, he configurado un controlador de dominio en una máquina virtual con Windows Server 2022, y he unido máquinas virtuales de clientes con Windows 11 Pro a dicho dominio.

Pasos para crear el **controlador de dominio**:

1. Habilitar HyperV en Windows Pro
2. Crear un conmutador externo “External Network”
3. Crear la VM para el Server con el conmutador externo.
4. Configurar una IP estática
 - Configurar una IP estática en la interfaz de red (192.168.1.20) esto se hace en Settings > Network & Internet > Ethernet.
 - Default gateway el de 8.8.8.8
 - Usar la IP del servidor para el DNS porque será el DNS del dominio (necesario para que los clientes encuentren el dominio)
5. Instalar Rol de Active Directory Domain Services
 - Abrimos Server Manager
 - Manage > Add Roles and Features
 - Server Selection -> Mi servidor local
 - Seleccionar en el rol el de Active Directory Domain Services
6. Promover el servidor a controlador de dominio
 - Desde el asistente en server manager, promover el servidor a domain controller y hacer toda la configuración del nombre de dominio raíz según el nombre del dominio que quiero dar, en este caso fue “cdserver.local”
 - Seleccionar la opción de Domain Name System y DHCP, esto es opcional pero yo he decidido realizarlo.

El server manager resultante de hacer toda esta serie de pasos se ve así:

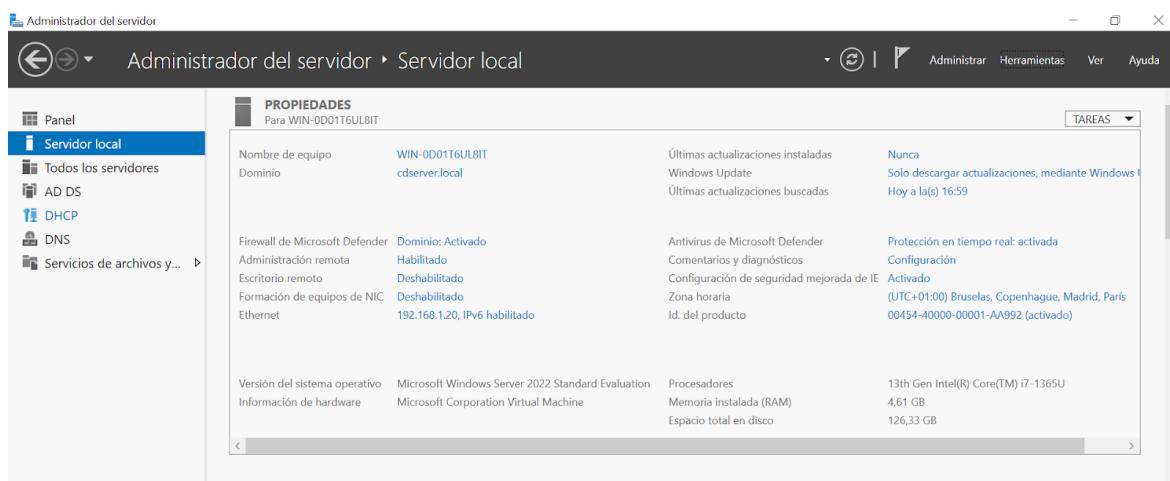


Figura 17. Aspecto servidor local

Pasos para configurar los **Cientes Windows**:

1. Crear la máquina virtual asignándole el mismo conmutador de red creado anteriormente.
2. Verificamos que el DNS primario apunta a la IP de mi controlador de dominio. Si no es así, se asigna manualmente.
3. Unimos la máquina al dominio definiendo en el nombre del equipo, el nombre del dominio:
 - Ingresamos a “Este equipo” “Propiedades” y cambiamos el nombre del equipo por el nombre de mi dominio.
 - Verificamos la acción introduciendo usuario y contraseña del administrador del dominio.

10.1.2 Implementación de usuarios híbridos

Una vez creada la red local, estamos preparados para realizar todo el proceso de configuración para poder obtener la identidad híbrida.

A continuación, detallo todos los pasos:

1. Se debe instalar Microsoft Entra Connect en el controlador de dominio:
 - Se debe tener la versión TLS 1.2, en caso de no tenerlo, se debe ejecutar el script “script_activar_tls.ps1” para poder actualizar la versión de TLS a la requerida.

```
PS C:\Users\Administrador\Desktop> echo "hola" > script_activar_tls_1.2.ps1
PS C:\Users\Administrador\Desktop> .\script_activar_tls_1.2.ps1
d TLS 1.2 has been enabled. You must restart the Windows Server for the changes to take affect.
PS C:\Users\Administrador\Desktop>
```

Figura 18. Habilitación TLS1.2

Posterior a la ejecución del script, se debe reiniciar el controlador de dominio.

2. Empezamos con la configuración:

A continuación, adjuntaré capturas de todo el proceso de configuración de Microsoft Entra Connect.

Primeramente, aceptamos términos y condiciones:

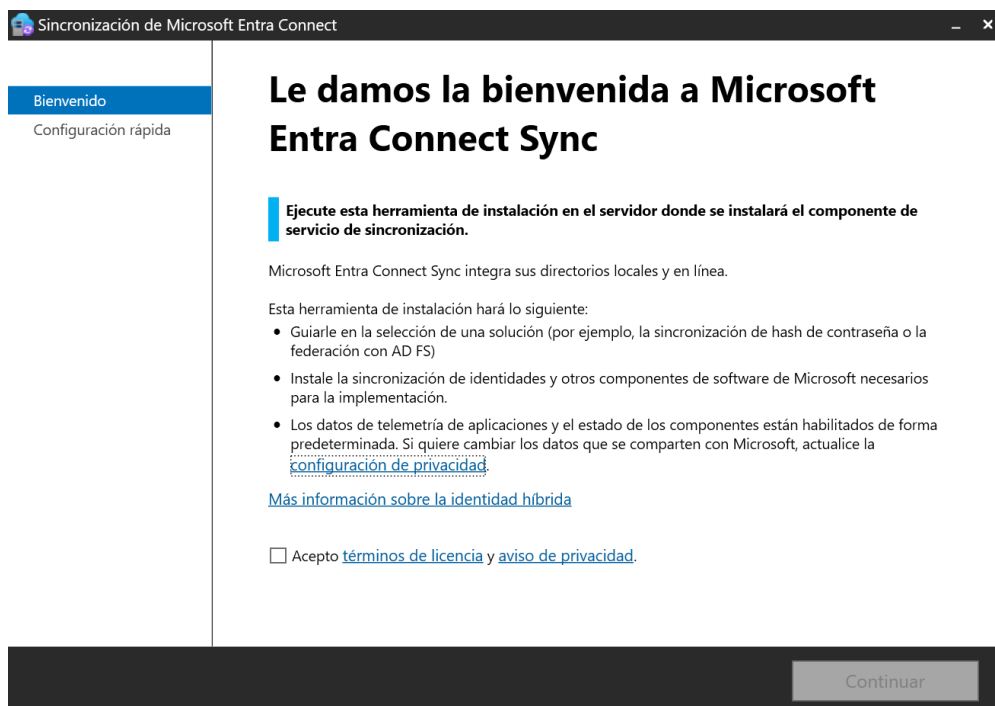


Figura 19. Bienvenida Microsoft Entra Connect

Ahora aparecerá una ventana en la que debemos escoger la configuración a realizar. Como se puede ver, se hace mención al bosque de Active Directory actual en el que se está realizando la configuración, y este es el mencionado anteriormente, “cdserver”

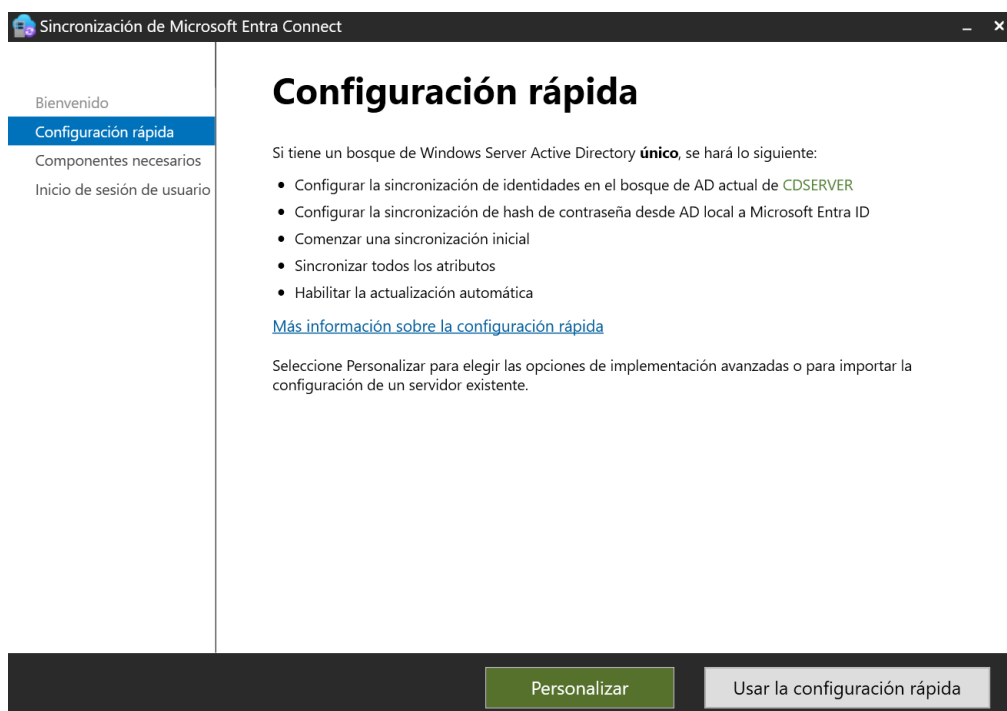
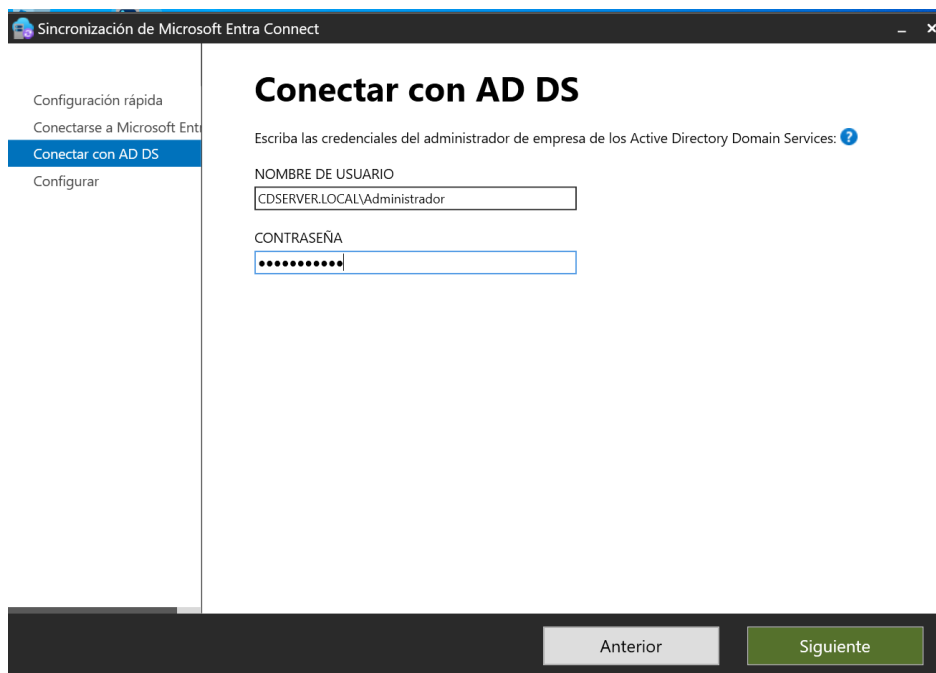


Figura 20. Selección de configuración Entra Connect

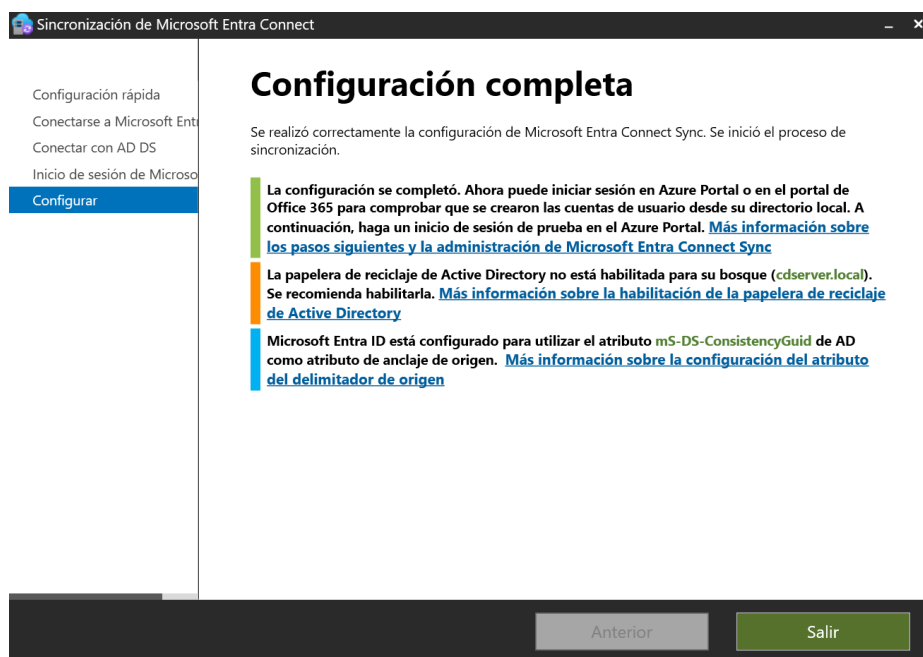
Posterior a ello aparecerá una pantalla donde se deben ingresar nuestras credenciales, deben ser las del administrador del controlador de dominio que queremos vincular:



The screenshot shows a window titled "Sincronización de Microsoft Entra Connect". On the left, a navigation pane lists: "Configuración rápida", "Conectarse a Microsoft Ent...", "Conectar con AD DS" (highlighted), and "Configurar". The main area is titled "Conectar con AD DS" and contains the instruction: "Escriba las credenciales del administrador de empresa de los Active Directory Domain Services: ?". Below this are two input fields: "NOMBRE DE USUARIO" with the text "CDSERVER.LOCAL\Administrador" and "CONTRASEÑA" with masked characters. At the bottom, there are two buttons: "Anterior" (disabled) and "Siguiente" (active).

Figura 21. Conectar AD DS

Una vez ingresadas las credenciales, se debe iniciar sesión en el tenant de Microsoft con el que se quiere establecer el vínculo. Una vez hecho esto, habremos completado la configuración, ya que aspectos más avanzados como configuración de UPN se ha decidido no realizarlo ya que es una simulación de una empresa más no es una empresa real y configurar UPN se saldría de nuestro objetivo actual.



The screenshot shows the same window as Figure 21, but now the navigation pane has "Configurar" highlighted. The main area is titled "Configuración completa" and contains the message: "Se realizó correctamente la configuración de Microsoft Entra Connect Sync. Se inició el proceso de sincronización." Below this are three informational blocks: a green one stating configuration is complete and providing links to Azure Portal and Office 365; an orange one stating that Active Directory recycling is not enabled for the "cdserver.local" forest and providing a link for more information; and a blue one stating that Microsoft Entra ID is configured to use the "mS-DS-ConsistencyGuid" attribute and providing a link for more information. At the bottom, there are two buttons: "Anterior" (disabled) and "Salir" (active).

Figura 22. Configuración completada

Una vez completada la configuración, obtuve un inicio de sesión único tanto en nuestras máquinas en local pertenecientes al dominio, como en todos los recursos en la nube de la empresa.

Todos los usuarios que había en el Active Directory del dominio ahora también están sincronizados en la nube de Azure bajo el mismo dominio. En este caso el dominio será el del tenant, pero si fuera una empresa real se hubiera configurado con la UPN que requiriera.

10.2 Fase de protección de datos

La fase de protección de datos abarca una serie de implementaciones de soluciones de Microsoft Purview. A continuación, detallaré la implementación que he tenido que seguir para poder desarrollarlas con éxito.

10.2.1 Sensitive Info Types

Como se mencionó anteriormente, los Sensitive Info Types son cruciales para poder definir la información que es sensible para una empresa. En el apartado de diseño he explicado qué decisiones he tomado para definir estos tipos, ahora explicaré con detalle cómo se configuraron para que estén correctamente definidas.

10.2.1.1 Product_Registration

Este Sensitive Info Type abarca todo lo relacionado a registros de patentes, marcas, modelos de utilidad y diseños industriales.

Para definirlo, he ingresado a Microsoft Purview, al apartado de Sensitive Info Types y he definido una lista de palabras clave:

Edit keyword list Clos

Keyword lists identify the words and phrases you want this info type to detect. For example, the keyword list to identify Netherlands VAT numbers is 'VAT number, vat no, vat number, VAT#'. [Learn how to create keyword lists](#)

☰ Choose from existing keyword lists

ID * ⓘ

Registry

Keyword group #1 * ⓘ 🗑️

Case insensitive

Marca nacional
SOLICITUD DE MODELO DE UTILIDAD
SOLICITUD DE PATENTE
Mark Information
RCD file information
EUTM file information

Figura 23. Patrón Registry

En esta pantalla final, se puede ver que se ha definido el patrón 1, con la lista de palabras clave “Registry” y la detección en cualquier parte del documento.

Define patterns for this sensitive info type

Sensitive info types are defined by one or more patterns. Each pattern must contain a primary element and confidence level, but you can also include supporting elements and additional checks to further refine the evaluation and detection of matching items. [Learn about defining patterns](#)




Name	Confidence level	
+ Create pattern 1 pattern <hr/> ^ Pattern #1 Primary element Keyword list: Registry Character proximity Detect primary AND supporting elements within 300 characters	High	  

Figura 24. Resumen patrón Registry

Estas palabras forman una lista de palabras clave que como su nombre indica, son cruciales en un registro de cualquiera de los documentos a proteger, son los títulos de los registros de cada trámite. Estas palabras están definidas en “Case insensitive” esto quiere decir que van a reconocer dichas palabras en cualquier parte de un documento o correo que las contenga, sea en mayúsculas o minúsculas. Esta misma estrategia de case insensitive fue la que se siguió en las demás definiciones de información sensible.

Entonces, si un agente está redactando un correo o documento, y contiene esas palabras, el sistema lo detectará como información sensible que luego será tratada según las etiquetas que tenemos definidas.

El próximo paso luego del resumen es confirmar la creación.

10.2.1.2 Invention_Proposal

Como se explicó anteriormente, este tipo de información sensible hace referencia a documentos o correos que contengan palabras claves referentes a contratos de confidencialidad o propuestas de invención.

Para poder implementar este tipo sensible, se han definido dos listas de palabras claves en Microsoft Purview:

Cada documento de propuesta de invención o acuerdo de confidencialidad está conformado por una cabecera con los títulos “Propuesta de invención” “Acuerdo de confidencialidad” en inglés o en español, seguido de unos campos rellenos con la información de cada cliente y proyecto. Es por ello por lo que se definieron dos patrones, uno para las cabeceras y otro para los campos.

Cada patrón forma una lista de palabras claves:

- Header:

Edit keyword list

no, vat number, VAT#. [Learn how to create keyword lists](#)

☰ Choose from existing keyword lists

ID * ⓘ

Header

Keyword group #1 * ⓘ

Case insensitive

Propuesta de invención
 Proposal for invention
 Confidentiality agreement
 Acuerdo de confidencialidad

Figura 25. Patrón Header

- Fields:

Keyword lists identify the words and phrases you want this info type to detect. For example, the keyword list to identify Netherlands VAT numbers is 'VAT number, vat no, vat number, VAT#'. [Learn how to create keyword lists](#)

☰ Choose from existing keyword lists

ID * ⓘ

Fields

Keyword group #1 * ⓘ

Case insensitive

Título:
 Autor:
 Descripción:
 Anexos:
 Titol:
 Autor:
 Description:

Figura 26. Patrón Fields

Ahora como se puede apreciar en la siguiente figura, se han añadido dos patrones, uno es el elemento primario, el cual está conformado por la cabecera y luego están los elementos secundarios, conformados por los campos. Además, cabe destacar que para poder clasificar esta información como sensible, deben aparecer ambos elementos en el mismo documento.

Define patterns for this sensitive info type

Sensitive info types are defined by one or more patterns. Each pattern must contain a primary element and confidence level, but you can also include supporting elements and additional checks to further refine the evaluation and detection of matching items. [Learn about defining patterns](#)

+ Create pattern 1 pattern




Name	Confidence level	
^ Pattern #1 Primary element Keyword list: Header Character proximity Detect primary AND supporting elements within unlimited characters Supporting elements Keyword list: Fields	High	  

Figura 27. Resumen patrón Header y Fields

10.2.1.3 Invoice


Como se explicó anteriormente en el apartado de diseño, este tipo de información sensible detecta patrones de divisas y facturas mediante tres patrones:

- **Pre_divisa_regex:** a diferencia de las palabras claves, este patrón es una expresión regular la cual detecta todo string que comience con una divisa delante, seguido de un número, por ejemplo: “\$50”, “€50” sería detectado. También sería detectado números del estilo “€1,000.50” o “\$1.000,50” cubriendo así ambas posibilidades y combinaciones de separadores. La expresión regular es la siguiente:


Add a regular expression

Regular expressions (RegEx) are strings of text that create patterns to help identify and match the info you're looking for. RegEx strings can be formatted many ways. For example `\d{6}` identifies a six-digit number in content. [Learn about supported RegEx patterns](#)

 Choose from existing regular expressions

ID * 

pre_divisa_regex

Regular expression * 

`((EUR|€|USD|¥)(\s?[0-9]*[.]?[0-9]*))`


Word match
 String match 

Figura 28. Regex pre_divisa

- **Post_divisa_regex:** detecta cuando un string comienza con un número y acaba con un símbolo de divisa, sea dólar o euro. Por ejemplo, “50€”, “50\$”, sería detectado, además de las diferentes combinaciones de separación en las cifras como “1.000,50€”, “1,000.50\$”:

Edit regular expression

Close

Regular expressions (RegEx) are strings of text that create patterns to help identify and match the info you're looking for. RegEx strings can be formatted many ways. For example `\d{6}` identifies a six-digit number in content. [Learn about supported RegEx patterns](#)

ID * ⓘ

final_divisa_regex

Regular expression * ⓘ

`[0-9]*[.]?[0-9]*[.]?[0-9]+\s?(€|EUROS|euros)[\s,,:]?`

Word match String match ⓘ

Figura 29. Regex final_divisa


- Invoice_list: en este patrón, volvemos a detectar listas de palabras claves ya que la finalidad es detectar títulos de documentos con facturas o la mención de estas palabras en correos. Al ser una empresa de abogados de la propiedad intelectual, he considerado muy importante proteger transacciones tales como facturas para mantener confidencial cualquier relación entre la empresa y los autores de ideas.

Add a keyword list

☰ Choose from existing keyword lists

ID * ⓘ

Invoice_list

Keyword group #1 * ⓘ 

Case insensitive

Invoice
Factura
Rechnung
Facture

Case sensitive

Enter keywords, separated by a new line. Each keyword is limited to 50 characters, and exact casing is required to detect matches.

Figura 30. Lista de palabras clave Invoice_list

En este caso, he definido los tres patrones como elementos primarios porque quiero que la detección de ellos sea independiente, ya que una práctica muy común es que se tenga

totalmente monitorizado este tema. Es decir, es una relación “OR” de elementos primarios, no hay elementos de soporte. El comportamiento para que el sistema detectara un documento con información sensible de este tipo sería que aparezca al menos uno de estos patrones.










+ Create pattern		3 patterns
Name	Confidence level	
^ Pattern #1 Primary element Regular expression: pre_divisa_regex Character proximity Detect primary AND supporting elements within 300 characters	High	  
^ Pattern #2 Primary element Regular expression: final_divisa_regex Character proximity Detect primary AND supporting elements within 300 characters	High	  
^ Pattern #3 Primary element Keyword list: Invoice_list Character proximity Detect primary AND supporting elements within 300 characters	High	  

Figura 31. Resumen de información sensible “Invoice”

10.2.2 Microsoft Information Protection

Se implementaron cuatro etiquetas de confidencialidad:

10.2.2.1 Confidential non encrypted

Esta etiqueta como su nombre indica, no está cifrada. Es decir, si un empleado clasifica sus documentos con esta etiqueta, cualquier persona podría leer su contenido, sea o no de la empresa. Los motivos para implementar esta etiqueta se explicaron anteriormente en el apartado de diseño.

La configuración se basa en lo siguiente:

- Un “Scope” de cualquier tipo de archivo, correo, Site de SharePoint o grupo de comunicación. Es decir, en todo el tenant de la empresa podrá utilizarse esta etiqueta.
- Access Control: Limita quién puede ver esta información etiquetada, se ha definido que no hay Access control.
- Se le aplica un footer con “Confidential – Unencrypted”
- El resto de opciones están desactivadas.

El resumen de la configuración es el siguiente:

Name
Confidential non encrypted
[Edit](#)

Display name
Confidential Unencrypted
[Edit](#)

Description for users
This label is designed to send unencrypted information to external recipients who require it.
[Edit](#)

Scope
Files & other data assets, Email, Site, UnifiedGroup
[Edit](#)

Access control
None
[Edit](#)

Content marking
Footer: Confidential - Unencrypted
[Edit](#)

Auto-labeling for files and emails
None
[Edit](#)

Group settings
None

Figura 32. Resumen de la etiqueta “Confidential non encrypted”

10.2.2.2 Confidencial Encrypted

Esta etiqueta cifra la información que la lleva. Se siguió la estrategia anteriormente explicada en el apartado de diseño de recomendación de etiqueta tras detectar la información sensible “IBAN, DNI, Número de pasaporte, Tarjetas débito/crédito e información sensible “Invoice”. Se siguió la configuración siguiente:

- **Access control:** se configuró que sólo tengan acceso los usuarios de la empresa

Access control

Use encryption capabilities to control who can access labeled items. Depending on the scope you specified, items can include emails, Office, Fabric and Power BI files, and meeting invites. [Learn more about access control settings](#)

Remove access control settings if already applied to items

Configure access control settings

Turn on co-authoring for Office desktop apps so multiple users can simultaneously edit labeled documents that have access control settings applied. [Learn more about this setting](#)

[Go to co-authoring setting](#)

Assign permissions now or let users decide?

Assign permissions now

The settings you choose will be automatically enforced when the label is applied to email and Office files.

User access to content expires ⓘ

Never

Allow offline access ⓘ

Never

Figura 33. Parte 1 control de acceso cifrado

Definimos que solo los usuarios con el dominio perteneciente al tenant de la empresa, tienen acceso:

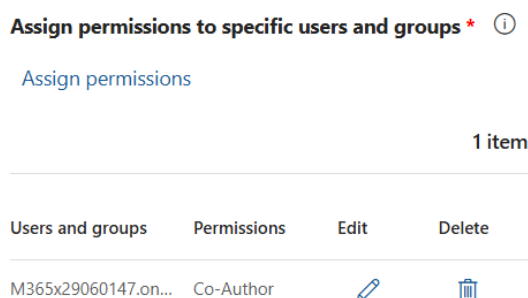


Figura 34. Parte 2 control de acceso cifrado

- **Content contains:** En este apartado, defino cuál es el contenido sensible a detectar para poder realizar la recomendación. Aquí se puede observar el grado de confidencialidad, el número de instancias a detectar para que sea un documento sensible y que puede estar en cualquier sitio del documento.

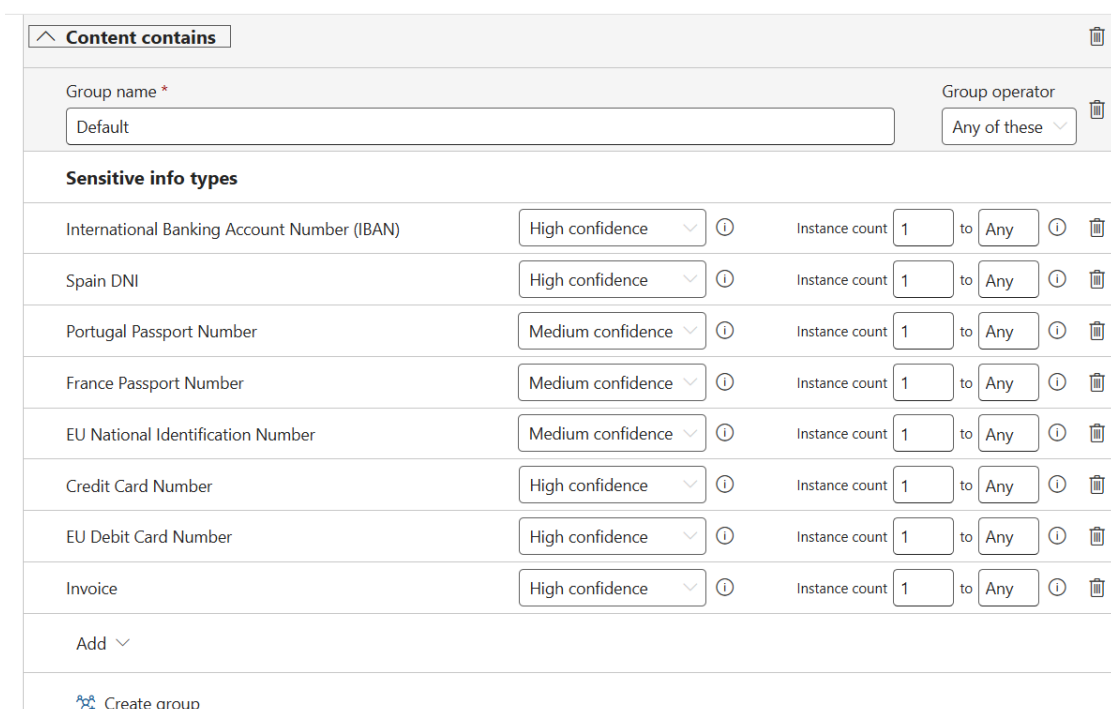


Figura 35. Definición de información sensible etiqueta “Confidential Encrypted”

- **Acción al detectar información sensible:** Esto es de las partes más importantes de la configuración de la etiqueta, y es que tras poner estas configuraciones conseguimos que el sistema recomiende al usuario aplicar la etiqueta confidencial debido a la información confidencial que está manejando.

When content matches these conditions

Recommend that users apply the label

Automatic and recommended labeling works differently for items in Office 365 vs. files stored on Windows devices. [Learn more](#)

Display this message to users when the label is applied ⓘ

Your organization recommends that you label this resource with the label Confidential Encrypted.

Figura 36. Definición de recomendación etiqueta “Confidential Encrypted”

10.2.2.3 Industrial Property Use

Esta etiqueta se implementó de manera que sólo tuvieran acceso los empleados pertenecientes al grupo “Intellectual Property Agent” y que la información que fuera detectada como sensible según la configuración realizada, se etiquetara automáticamente. La configuración fue la siguiente:

- Access control: sólo pueden leer la información los usuarios pertenecientes al grupo “Intellectual Property Agent”.

Assign permissions now or let users decide?

Assign permissions now

The settings you choose will be automatically enforced when the

User access to content expires ⓘ

Never

Allow offline access ⓘ

Always

Assign permissions to specific users and groups * ⓘ

Assign permissions

Users and groups

Use dynamic watermarking ⓘ

Use Double Key Encryption ⓘ

Back Next

Assign permissions

Only the users or groups you choose will be assigned permissions to use the content that has this label applied. You can choose from existing permissions (such as Co-Owner, Co-Author, and Reviewer) or customize them to meet your needs.

- + Add all users and groups in your organization
- + Add users or groups **ated users** ⓘ
- + Add users or groups
- + Add specific email addresses or domains ⓘ

1 item

Permissions assigned to	Delete
IntellectualPropertyAgent@M365x29060147.onmicrosoft.com	

Choose permissions

Editor

View content,View rights,Edit content,Save,Print,Copy and extract content,Reply,Reply all,Forward,Allow macros

Save Cancel

Figura 37. Detalle del control de acceso.

- Opción auto-labeling: Está habilitada y es la que permite la funcionalidad de autoclasificar documentos.

Auto-labeling for files and emails

When users edit Office files or compose, reply to, or forward emails from Outlook that contain content matching the conditions you choose here, we'll automatically apply this label or recommend that they apply it themselves. [Learn more about auto-labeling for Microsoft Purview](#)

To automatically apply this label to files that are already saved (in SharePoint and OneDrive) or emails that are already processed by Exchange, you must create an auto-labeling policy. [Learn more about auto-labeling policies](#)

Auto-labeling for files and emails



Since encryption is turned on, a large amount of content might be automatically encrypted when this label is applied. Turning on encryption impacts Office files (Word, PowerPoint, Excel) and PDF files that have this label applied. Because the files will be encrypted for security reasons, performance will be slow when the files are opened or saved, and some SharePoint and OneDrive features will be limited or unavailable. [Learn more](#)

^ Detect content that matches these conditions

Figura 38. Auto-labelig etiqueta Industrial Property Use

- **Content contains:** Aquí se define la información que el sistema detectará. Cuando el sistema la detecte, se autoclasificará el documento, restringiendo así el acceso a los usuarios fuera del grupo anteriormente mencionado. Vemos que se define como información sensible todo lo definido anteriormente referente a contratos de confidencialidad, facturas y registros realizados.

^ Detect content that matches these conditions

^ **Content contains**
🗑️

Group name *

Group operator

Any of these
🗑️

Sensitive info types

Invention_Proposal	High confidence	Instance count	1	to	Any	🗑️
Product_Registration	High confidence	Instance count	1	to	Any	🗑️

Add ▾

[🔗 Create group](#)

+ Add condition ▾

When content matches these conditions

Automatically apply the label
▾

Automatic and recommended labeling works differently for items in Office 365 vs. files stored on Windows devices. [Learn more](#)

Figura 39. Definición información sensible etiqueta Industrial Property Use

10.2.2.4 Internal Use

Esta etiqueta está configurada de manera que sea la etiqueta por defecto de la empresa. Tiene control de acceso, es decir, está cifrada y solo puede verla todos los usuarios pertenecientes al tenant de la empresa. La configuración es la siguiente:

Review your settings and finish

Name
Internal Use

Display name
Internal Use
[Edit](#)

Description for users
This label is for internal-use resources that are not highly confidential.
[Edit](#)

Scope
Files & other data assets, Email, Meetings, Site, UnifiedGroup
[Edit](#)

Access control
Access control
[Edit](#)

Content marking
Footer: Internal Use
[Edit](#)

Auto-labeling for files and emails
None
[Edit](#)

Group settings
None
[Edit](#)

Figura 40. Configuración etiqueta Internal Use

Para finalizar con este apartado de Microsoft Information Protection, cabe destacar que cada etiqueta es creada, pero no están publicadas hasta que son publicadas con una política de publicación.

Por ello, enseñaré como ejemplo la configuración de política de publicación de la etiqueta Confidential Encrypted, pero con todas las etiquetas he seguido la misma configuración:

Policy for confidential encrypted info

[Edit policy](#) [Delete policy](#)

Name
Policy for confidential encrypted info

Description

Published labels
Confidential Encrypted

Admin units
None

Published to
Exchange email - All accounts

Policy settings
Users must provide justification to remove a label or lower its classification

Figura 41. Configuración de políticas de publicación de las etiquetas

Lo destacable es que está publicado para todas las cuentas del tenant y que los usuarios deben proporcionar una justificación para remover la etiqueta o bajar su clasificación.

10.2.3 Data loss Prevention

Se ha implementado una política DLP para realizar una acción tras reconocer información sensible para la empresa.

10.2.3.1 DLP Unencrypted Label

Para poder cumplir con el cometido de realizar una acción de pedir justificación para la compartición a usuarios externos a la empresa, se ha tenido que seguir la siguiente configuración:

- **Rule DLP Label Unencrypted:** Esta regla está basada en condiciones AND. Cuando detecte que está utilizándose la etiqueta “Confidential Unencrypted” y el contenido es compartido para usuarios externos a mi organización la política lo que hará es mostrar un “Policy Tip” avisando sobre la alerta de la política con un enlace a “Override” el cual mostrará un formulario referente a justificar el motivo por el que está compartiendo información con o sin cifrar a un externo.

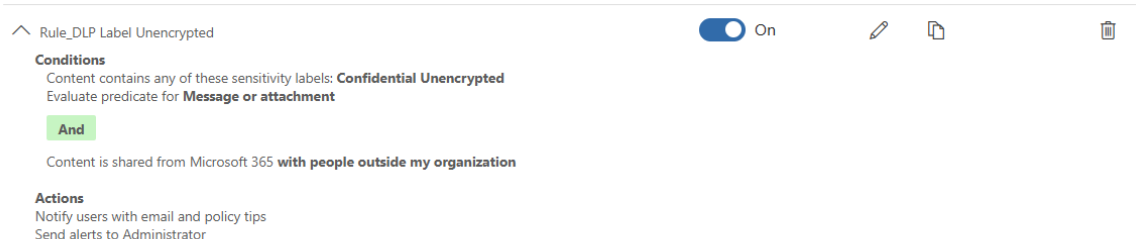


Figura 42. Configuración resumen DLP

Para que sea posible esta configuración he tenido que habilitar diversos campos en la regla:

- **User notifications:**

Este apartado es sobre informar al usuario y educarlo sobre la gravedad de la información sensible que comparte.

Por una parte, se habilita recibir notificaciones de email con aviso sobre la vulneración de la política y, por otra parte, se habilita el “Policy tip”, el cual es el encargado de explicar cuál es la política que se está vulnerando y las acciones que debe llevar a cabo el empleado.

Para mayor claridad, en la figura 43 podemos ver una ilustración de lo que es el “Policy tip” y cómo se mostraría al usuario al cumplir con las condiciones. En este caso se comparte a un externo y además se aplicó la etiqueta de sensibilidad sin cifrado.

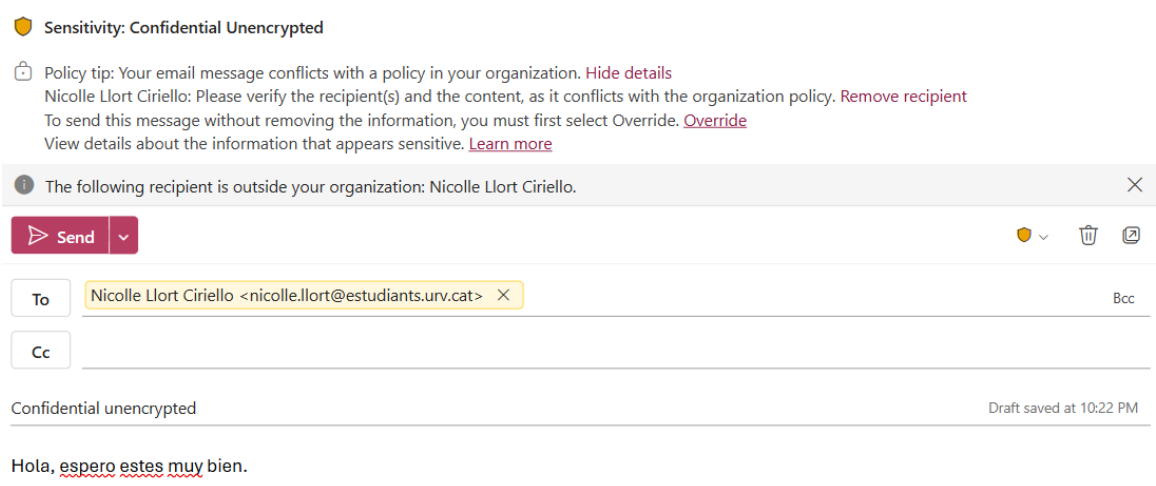


Figura 43. Demostración Policy tip

Otra parte importante de “User Notifications” es el “Override. Como se puede observar en la figura 43, aparece el policy tip avisando al usuario que su mensaje entra en conflicto con una política de la organización y más abajo se indica que para poder enviar este mensaje se debe seleccionar “Override”.

Override es una opción que permite a los usuarios sobrescribir políticas, es decir, omitirlas bajo una serie de condiciones configuradas. En este caso se ha configurado que al enviar correos a un externo y al enviar mensajes sin cifrado, se pueda omitir la política y realizar el envío siempre y cuando proporciones una justificación.

Si el usuario desea continuar con su envío bajo estas condiciones, debe presionar obligatoriamente “Override”. Al hacer esto se presentará el siguiente pop-up frente al usuario.

Confirmar la invalidación

⚠ **Sugerencia de directiva:** This item conflicts with a policy in your organization. [Mostrar detalles](#)

Su correo electrónico mensaje está en conflicto con una directiva de su organización. Ha intentado enviar contenido clasificado como Confidential Unencrypted a uno o varios destinatarios externos, considere la posibilidad de quitar estos destinatarios o reemplazar el contenido.

Justificación

Este destinatario tiene derecho a recibir este contenido

Mi administrador ha aprobado el uso compartido de este contenido

Otra justificación

Explique aquí su justificación empresarial.

Reconocimiento

Confirmando que esta comunicación cumple con las directivas y estándares de mi organización.

Figura 44. Demostración Override.

Una vez explicadas las funcionalidades y conceptos, se muestra la configuración total seguida en User notifications:

^ **User notifications**

Use notifications to inform your users and help educate them on the proper use of sensitive info.

On

Email notifications

[Preview and edit notification email](#)

Notify the user who sent, shared, or last modified the content.

Notify these people:

Attach matching email message to the notification (applies only to Exchange)

Policy tips

Support and behavior for policy tips varies across apps and platforms. [Learn where policy tips are supported](#)

Customize the policy tip text

Show the policy tip as a dialog for the end user before send (available for Exchange workload only)

To help ensure all email messages display the pop-up before they're sent, you must first configure Group Policy Object (GPO) settings to allow for full e

Upload a JSON file containing custom content that will be used in the pop-up dialog ⓘ

Provide a compliance URL for the end user to learn more about your organization's policies (available for Exchange workload only)

Figura 45. Configuración DLP, notificaciones email y policy tip.

Por último, adjunto la configuración para habilitar el Override:

^ **User overrides**

Allow overrides from Microsoft 365 files and Microsoft Fabric items

Allow users to override policy restrictions in Fabric (including Power BI), Exchange, SharePoint, OneDrive, and Teams.

Require a business justification to override

Override the rule automatically if they report it as a false positive

Require the end user to explicitly acknowledge the override (available for Exchange workload only)

Figura 46. Configuración DLP, notificaciones email y policy tip.

10.2.4 Insider Risk Management

10.2.4.1 Requisitos previos

Las configuraciones de IRM que he realizado requieren de tener los dispositivos de los empleados dados de alta en el portal de Microsoft Purview y vinculados a un correo de un empleado. Es por ello, que antes de configurar las políticas de IRM, para que estas puedan funcionar se debe hacer la configuración de alta y vinculación de dispositivos.

10.2.4.1.1 Alta de dispositivos

Para dar de alta un dispositivo se debe instalar el agente de Microsoft Purview indicado para ello y ejecutarlo:

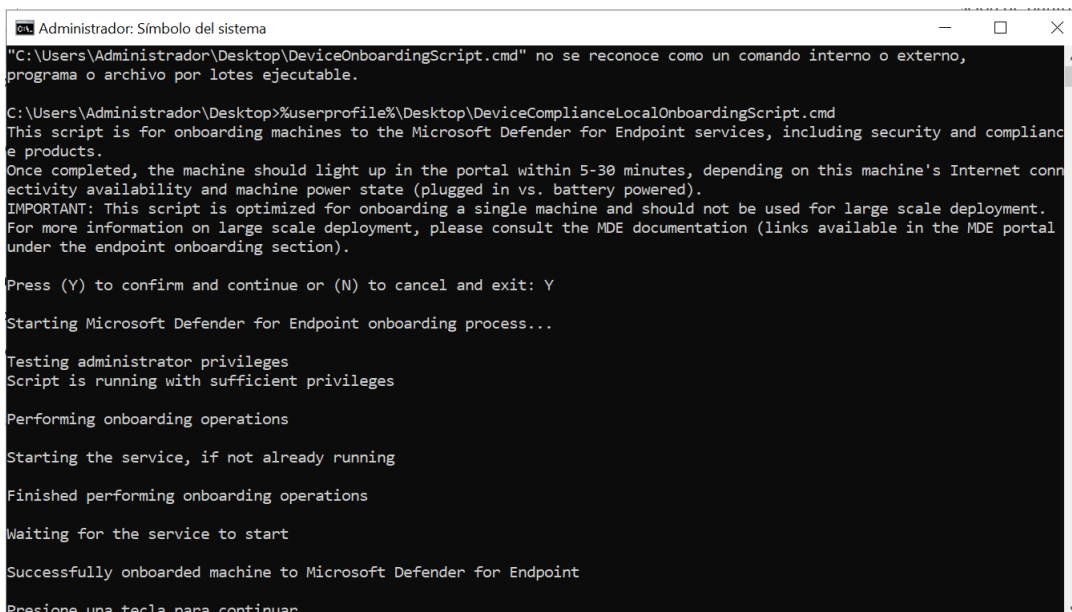


Figura 47. Instalación de agente para vincular dispositivos

Con el motivo de realizar una prueba integral de la solución, he realizado el Onboarding de tres dispositivos los cuales son tres máquinas virtuales que he unido al dominio previamente. Estas tres máquinas virtuales simulan tres empleados que también cuentan con la identidad híbrida SSO.

Una vez ejecutado el agente de Microsoft, aparecerán estos dispositivos dados de alta en el gestor de identidad de Azure, Entra ID:

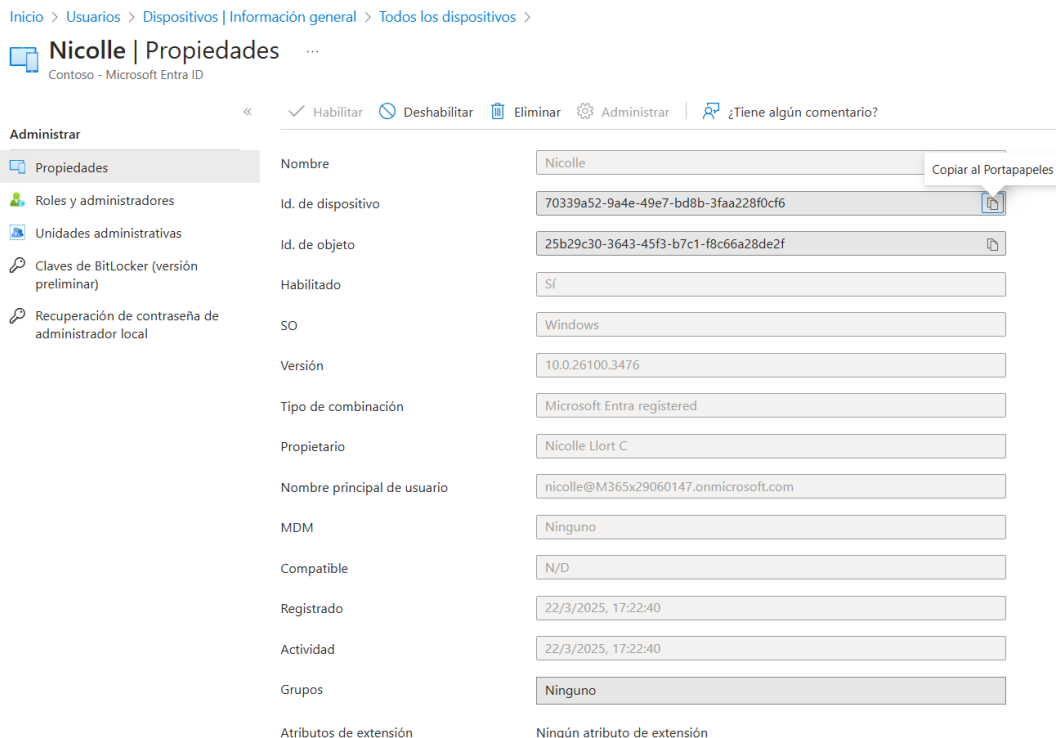


Figura 48. Dispositivo 1 dado de alta en Microsoft Entra ID

Inicio > Usuarios > **Dispositivos | Información general** > Todos los dispositivos >

Alan | Propiedades ...
Contoso - Microsoft Entra ID

« Habilitar Deshabilitar Eliminar Administrar | ¿Tiene algún comentario?

Administrar

- Propiedades
- Roles y administradores
- Unidades administrativas
- Claves de BitLocker (versión preliminar)
- Recuperación de contraseña de administrador local

Nombre	Alan	<input type="button" value="Copiar al Portapape"/>
Id. de dispositivo	d370d7ff-cb8a-4ea8-8a65-492de2f40379	<input type="button" value="Copiar"/>
Id. de objeto	367edcb1-b808-4960-b62d-5031973901d6	<input type="button" value="Copiar"/>
Habilitado	<input checked="" type="checkbox"/> Sí	
SO	Windows	
Versión	10.0.26100.3476	
Tipo de combinación	Microsoft Entra registered	
Propietario	María Perez	
Nombre principal de usuario	mariaperez@M365x29060147.onmicrosoft.com	
MDM	Ninguno	
Compatible	N/D	
Registrado	26/3/2025, 22:18:33	
Actividad	26/3/2025, 22:18:33	
Grupos	Ninguno	
Atributos de extensión	Ningún atributo de extensión	

[vices/DevicesMenuBlade/~/Overview/...](#)

Figura 49. Dispositivo 2 dado de alta en Microsoft Entra ID

Inicio > Usuarios > **Dispositivos | Información general** > Todos los dispositivos >

Ada | Propiedades ...
Contoso - Microsoft Entra ID

« Habilitar Deshabilitar Eliminar Administrar | ¿Tiene algún comentario?

Administrar

- Propiedades
- Roles y administradores
- Unidades administrativas
- Claves de BitLocker (versión preliminar)
- Recuperación de contraseña de administrador local

Nombre	Ada	<input type="button" value="Copiar al Portapape"/>
Id. de dispositivo	81ce94b9-fcd4-4bbd-995a-877530db2e28	<input type="button" value="Copiar"/>
Id. de objeto	aeb98e25-ff8c-49e2-ac2e-8744dfaae89b	<input type="button" value="Copiar"/>
Habilitado	<input checked="" type="checkbox"/> Sí	
SO	Windows	
Versión	10.0.26100.3476	
Tipo de combinación	Microsoft Entra registered	
Propietario	Jorge Rojas	
Nombre principal de usuario	jorgerojas@M365x29060147.onmicrosoft.com	
MDM	Ninguno	
Compatible	N/D	
Registrado	22/3/2025, 18:07:33	
Actividad	22/3/2025, 18:07:33	
Grupos	Ninguno	
Atributos de extensión	Ningún atributo de extensión	

Figura 50. Dispositivo 3 dado de alta en Microsoft Entra ID

Así como también aparecerán dichos dispositivos en el apartado de dispositivos dados de alta en Microsoft Purview:

Devices

Activity from these onboarded devices can be reviewed in features such as activity explorer or monitored by compliance solutions such as insider risk management and data loss prevention (DLP). [Learn more](#)

ⓘ Configuration or policy sync status is not available for some devices. There are a couple actions you can take to view these device details that aren't yet available:

- Update OS version for the device. Do this if the current OS version doesn't meet the minimum version for [Windows](#) and [MacOS](#) devices.
- Wait for device details to appear. If the device was recently onboarded, wait for the newly onboarded machine to gather the device details.

↓ Export Turn off Windows device monitoring Turn off macOS device monitoring

Filter

Days look back: 30 Configuration status: Any Policy sync status: Any OS: Any OS version: Any First onboarded date: Any Valid user:

Device name	Configuration status	Policy sync status	Last seen	OS
alan.cdserver.local	✔ Updated	✔ Updated	Mar 28, 2025 2:27 PM	Windows11
ada.cdserver.local	✔ Updated	✔ Updated	Mar 28, 2025 2:26 PM	Windows11
nicolle.cdserver.local	✔ Updated	✔ Updated	Mar 28, 2025 2:10 PM	Windows11

Figura 51. Dispositivos dados de alta en Microsoft Purview

- Vinculación de dispositivo a usuario

Una vez dado de alta los dispositivos que ahora son gestionados por el tenant, para poder detectar información de un usuario debemos vincularlo a uno de los dispositivos que hemos dado de alta. Para ello, se debe hacer lo siguiente:

1. Se debe descargar la aplicación “Portal de empleados” e iniciar sesión con el usuario en Microsoft Intune, una vez hecho esto aparecerá la ventana de confirmación:

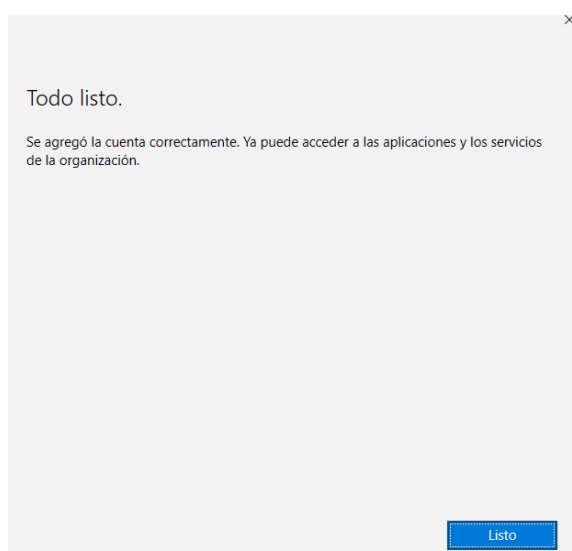


Figura 52. Pantalla confirmación Intune

Una vez hecha la configuración de Intune, vemos que efectivamente mariaperez@m365x29060147.onmicrosoft.com tiene asociado un dispositivo en Entra ID:

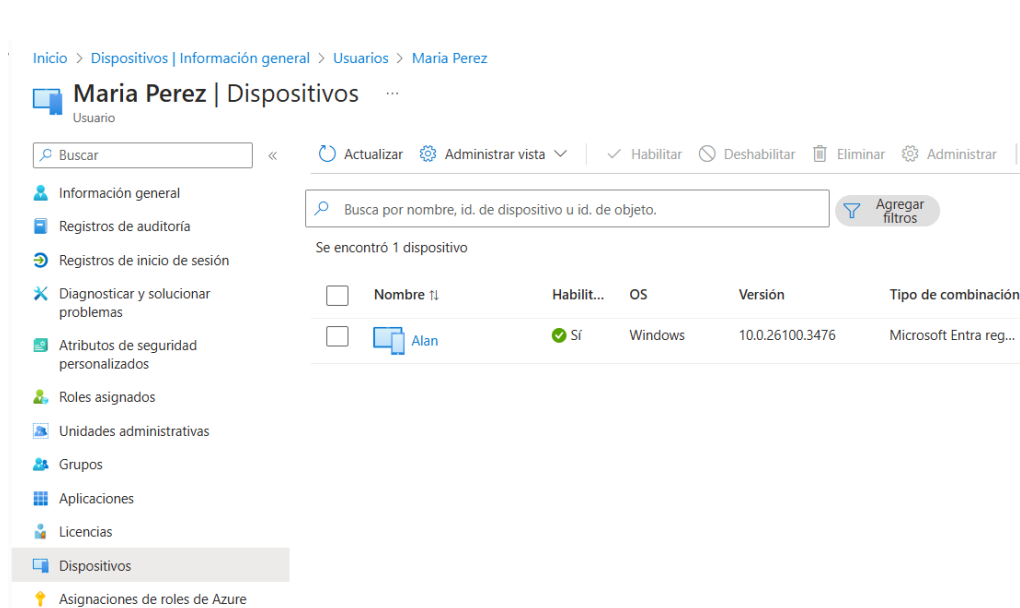


Figura 53. Vinculación usuario con dispositivo

Una vez confirmado esto, realicé el mismo procedimiento para todos los demás usuarios que queremos dar de alta.

10.2.4.1.2 Insider Risk Management Settings

Otro requisito previo para usar esta herramienta es configurar los dominios de confianza, de riesgo y habilitar o no la opción de anonimato.

- **Anonimato**

Esta configuración permite al administrador de alertas decidir si el nombre del usuario que genera la alerta será visible o permanecerá oculto. Esta opción resulta especialmente útil para garantizar la imparcialidad en la evaluación de alertas, ya que evita que el evaluador se vea influenciado, de manera directa o indirecta, por la identidad del autor de la alerta.

Por ejemplo, podría ocurrir que el administrador de alertas conozca y tenga una relación de amistad con el autor de la alerta, lo cual podría influir en su criterio y llevar a una evaluación menos rigurosa. Por otro lado, también es posible el caso opuesto, en el que se realice una evaluación más estricta debido a la percepción personal del evaluador. En ambos escenarios, esta ambigüedad puede ser eliminada al seleccionar una configuración anonimizada. De este modo, la persona encargada de evaluar no sabe quién generó la alerta, asegurando que todos los usuarios sean evaluados de manera equitativa.

Además, esta configuración de anonimato es reversible. Una vez completada la evaluación de las alertas, se puede desactivar para revelar la identidad de los autores, si es necesario.

Como demostración, he optado por mantener activado el anonimato en una de las alertas generadas de forma temporal, mientras que en la otra he dejado visibles los nombres desde el inicio.

Insider Risk Management settings

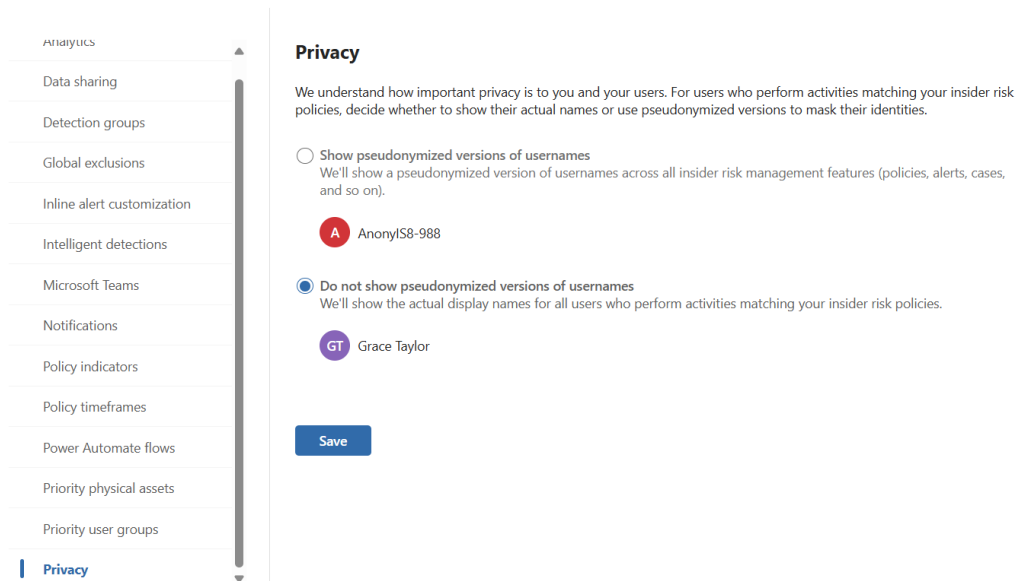


Figura 54. IRM Settings - Anonimato

- **Grupos de detección de dominios**

Un grupo de detección permite delimitar un indicador o una exclusión global para enfocarse en actividades de alto valor que son relevantes para la organización. En mi caso, este grupo lo he utilizado para establecer una lista de dominios confiables. Estos dominios pueden pertenecer a proveedores de la organización o ser aquellos que los empleados utilizan habitualmente en sus tareas laborales.

Posteriormente, esta lista será integrada en la configuración de exclusiones globales. Los dominios incluidos son los siguientes:

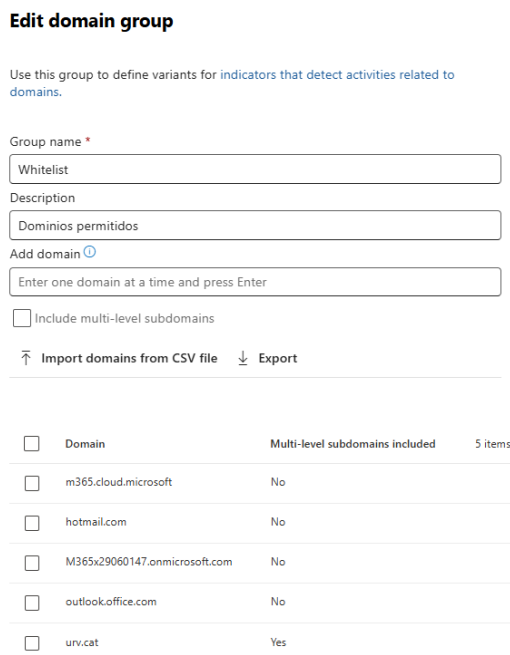


Figura 55. IRM Settings - Whitelist

Es importante destacar que la creación de esta lista dependerá completamente de la dimensión de la organización y de su capacidad para que el personal evalúe las alertas generadas. Por ejemplo, si una empresa utiliza el paquete Office y lo incluye en la lista de dominios confiables, no se detectarán alertas relacionadas con inicios de sesión y cargas de archivos en cuentas de Office que no pertenezcan a la organización.

Este riesgo puede considerarse aceptable en empresas de gran tamaño, con miles de empleados, donde sería inviable evaluar millones de alertas diariamente, ya que la mayoría de estas serían falsos positivos. Dado que este es el caso más común, he decidido implementar la lista de esta manera.

En cambio, para organizaciones más pequeñas, que generan pocas alertas al día, es posible considerar una mayor restricción de esta lista, eliminando los dominios relacionados con paquetes Office y Exchange para mejorar la precisión de las alertas.

- **Exclusiones globales**

En este apartado he añadido la lista de dominios confiables explicada anteriormente, esto con la finalidad de que la actividad referente a estos dominios sea ignorada, ya que no se considera de riesgo.

- **Intelligent Detections**

Se trata de configuraciones avanzadas que permiten al sistema identificar cuándo una alerta representa un riesgo más elevado en comparación con otras. A través de estas configuraciones, he definido una lista de dominios de riesgo elevado. Si estos dominios son detectados en una alerta, el sistema incrementará su nivel de riesgo por encima del estimado para la cantidad habitual de alertas generadas.

Por ejemplo, mientras que la detección de dos alertas en un dominio genérico podría considerarse de riesgo medio, la detección de dos alertas en dominios incluidos en esta lista aumentará automáticamente su nivel de riesgo a alto.

La lista es la siguiente:

+ Add domains		↑ Import domains from CSV file	↓ Export	8 items
<input type="checkbox"/>	Domain		Multi-level subdomains included	
<input type="checkbox"/>	drive.google.com		No	
<input type="checkbox"/>	web.whatsapp.com		No	
<input type="checkbox"/>	ilovepdf.com		No	
<input type="checkbox"/>	live.com		No	
<input type="checkbox"/>	gmail.com		No	
<input type="checkbox"/>	163.com		No	
<input type="checkbox"/>	google.mail.com		No	
<input type="checkbox"/>	docs.google.com		No	

Figura 56 IRM Settings – Dominios alto riesgo

10.2.4.2 Indicadores de las políticas implementadas

Los indicadores son la actividad que se utiliza para identificar comportamientos o actividades que podrían representar un riesgo interno para la organización. Estos, ayudan a monitorizar y detectar actividades de empleados perjudiciales para la empresa. Se han implementado cuatro indicadores:

- **Sending email with attachments**

Este indicador monitorea el envío de correos electrónicos que contienen archivos adjuntos, esto es útil para identificar posibles riesgos relacionados a la transferencia de información sensible fuera de la organización.

- **Using a browser to upload files to the web**

Este indicador monitorea las actividades relacionadas a la carga de archivos desde un navegador web a una plataforma en internet. Esto es muy útil para detectar posibles riesgos de exfiltración de datos o transferencias no autorizadas de información confidencial a servicios externos. Un ejemplo de ello puede ser subir información sensible a aplicaciones web de análisis de texto, resúmenes o conversores de ficheros poco fiables.

Para que este indicador funcione es obligatorio registrar los dispositivos que utilizan los usuarios a monitorizar, ya que el registro de dispositivos permite capturar en tiempo real todos los datos relacionados a las conexiones a internet, usando en este caso el navegador web.

- **Entering risky prompt in Copilot**

Este indicador detecta cuando se envía un mensaje con información sensible a Copilot. Es muy interesante ya que detecta mensajes tanto escritos como adjuntados, permitiendo un monitoreo más extenso.

- **Receiving sensitive response from Copilot**

Este indicador detecta cuando se recibe una respuesta con información sensible proveniente de Copilot.

10.2.4.3 Políticas de Insider Risk Management implementadas

En la siguiente sección detallaremos las políticas implementadas.

10.2.4.3.1 File Uploads

Esta política se ha implementado para detectar las subidas de archivos a dominios externos realizadas por los empleados, incluyendo también la actividad relacionada con introducción de prompts y respuestas de riesgo a Copilot.

Así como Intelligent Detections incrementa el nivel de riesgo en una alerta, añadir los Sensitive Info Types en la captura siguiente, me permite priorizar ese contenido que la organización considera sensible y aumentar el nivel de riesgo si lo detecta:

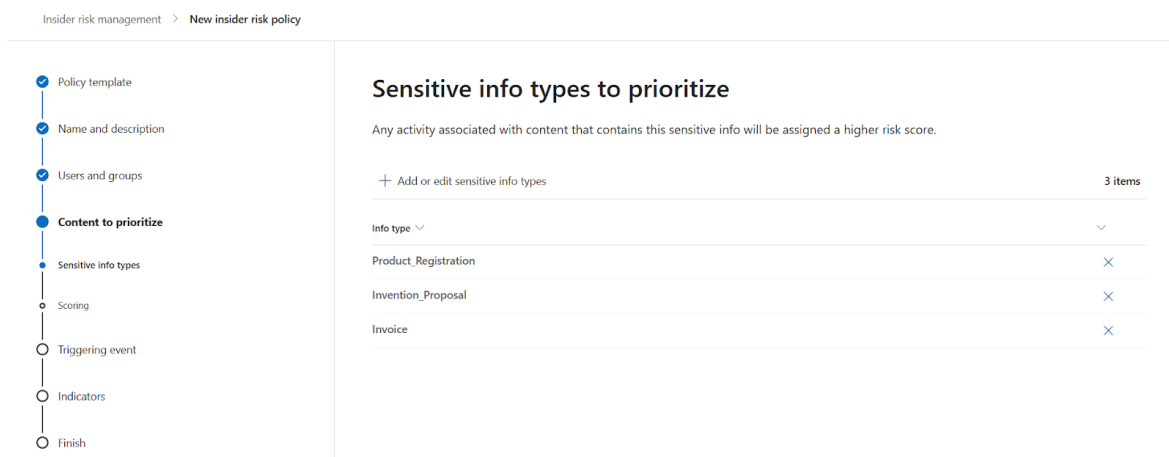


Figura 57. IRM File Uploads – Contenido prioritario

Sin embargo, seguirá detectando todas las alertas que se ocasionen bajo el caso de estudio, no solo las que contengan información sensible, como puede verse en la imagen siguiente:

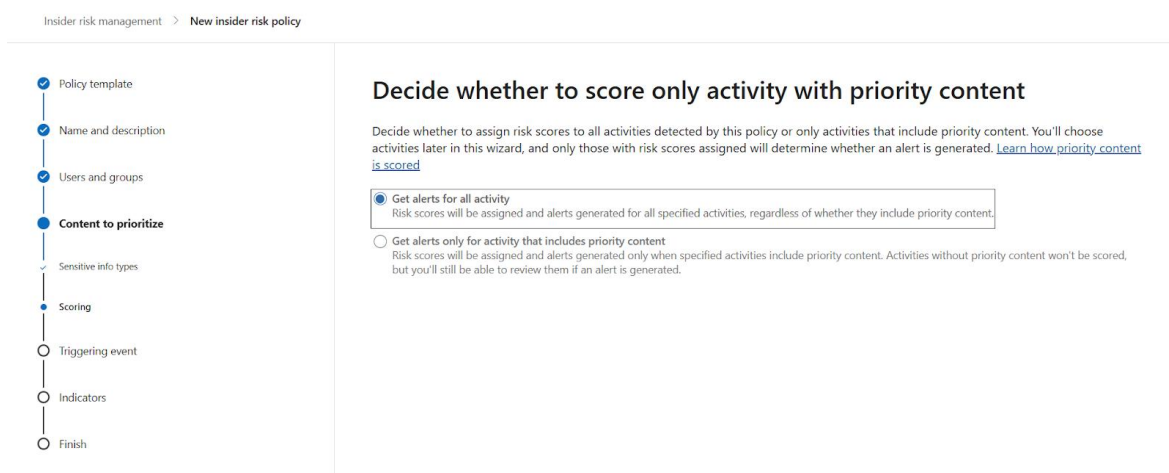


Figura 58. IRM File Uploads – Detección de toda la actividad generada

Las políticas de Insider Risk Management necesitan detectar una serie de secuencias o eventos para comenzar a estudiar a un usuario. En este caso seleccioné las actividades de:

- Usar un navegador para subir documentos a la web
- Degradar o remover una etiqueta y luego exfiltrar
- Degradar o remover una etiqueta, descargarla y luego exfiltrarla
- Degradar o remover una etiqueta, exfiltrarla y luego eliminarla
- Degradar o remover una etiqueta, ofuscar y luego exfiltrar

Se puede ver cada uno de esos eventos en la captura de configuración siguiente:

Choose one or more triggering events to determine when a policy will begin assigning risk scores to a user's activity.[Learn more](#)

User matches a data loss prevention (DLP) policy
 Policy will start assigning risk scores when a user performs an activity matching the DLP policy you select. The DLP policy must be configured to generate 'High' severity incident reports. [Learn more about DLP policy requirements](#).
 You do not have any DLP policies configured. Without a DLP policy selected, this policy will not assign risk scores to user's activity. [Learn more](#)

User performs an exfiltration activity
 Policy will start assigning risk scores when specific thresholds are detected for activity relating to the following indicators:
Select which activities will trigger this policy

- Downloading content from SharePoint
- Sending email with attachments to recipients outside the organization
- Printing files
- Creating or copying files to USB
- Using a browser to upload files to the web
- Sharing SharePoint files with people outside the organization
- Archive, obfuscate, exfiltrate, then delete
- Downgrade or remove label then exfiltrate
- Downgrade or remove label, download, then exfiltrate
- Downgrade or remove label, download, exfiltrate, then delete
- Downgrade or remove label, download, obfuscate, then exfiltrate
- Download from third-party site then exfiltrate
- Download from third-party site, exfiltrate, then delete
- Download from third-party site, obfuscate, exfiltrate, then delete

Figura 59. IRM File Uploads – Eventos desencadenantes

Estos eventos anteriormente explicados empezarán a ser estudiados cuando se supere el umbral que he definido para cada uno de ellos. El umbral definido es igual para todos: tres eventos en total al día, un evento de archivos con información sensible, un evento con archivos de contenido prioritario y uno de subidas a dominios no permitidos.

A continuación, muestro la configuración de umbrales realizada para cada evento desencadenante.

- Using a browser to upload files to the web será estudiado a partir de los siguientes umbrales:

Choose thresholds for triggering events

The policy will start assigning risk scores to activity only when specific thresholds are met for the exfiltration activities you selected as the triggering event. Thresholds are based on the number of events recorded for an activity per day. You can use recommended thresholds or specify your own.

Apply built-in thresholds RECOMMENDED

Choose your own thresholds

Using a browser to upload files to the web

- Total number of activities
 per day
- Number of activities for files containing sensitive info types
 per day
- Number of activities for files matching priority content
 per day
- Number of activities performed in which target is unallowed domain
 per day

Figura 60. IRM File Uploads – Umbrales de eventos 1

- Los eventos de degradar o remover una etiqueta y luego exfiltrar serán estudiados a partir de los siguientes umbrales:

Downgrade or remove label then exfiltrate

Total number of activities
3 per day

Number of activities for files containing sensitive info types
1 per day

Number of activities for files matching priority content
1 per day

Number of activities performed in which target is unallowed domain
1 per day

Activity is above user's usual activity for the day

[Reset to defaults](#)

Figura 61. IRM File Uploads – Umbrales de eventos 2

- Degradar o remover una etiqueta, descargarla y luego exfiltrarla, serán eventos estudiados a partir de los siguientes umbrales:

Downgrade or remove label, download, then exfiltrate

Total number of activities
3 per day

Number of activities for files containing sensitive info types
1 per day

Number of activities for files matching priority content
1 per day

Number of activities performed in which target is unallowed domain
1 per day

Activity is above user's usual activity for the day

[Reset to defaults](#)

Figura 62. IRM File Uploads – Umbrales de eventos 3

- Degradar o remover una etiqueta, exfiltrarla y luego eliminarla, es una secuencia de eventos que será estudiada a partir de los siguientes umbrales:

Downgrade or remove label, download, exfiltrate, then delete

Total number of activities
 per day

Number of activities for files containing sensitive info types
 per day

Number of activities for files matching priority content
 per day

Number of activities performed in which target is unallowed domain
 per day

Activity is above user's usual activity for the day

[Reset to defaults](#)

Figura 63. IRM File Uploads – Umbrales de eventos 4

- Degradar o remover una etiqueta, descargar, ofuscar y luego exfiltrar es una secuencia que será estudiada a partir de los siguientes umbrales:

Downgrade or remove label, download, obfuscate, then exfiltrate

Total number of activities
 per day

Number of activities for files containing sensitive info types
 per day

Number of activities for files matching priority content
 per day

Number of activities performed in which target is unallowed domain
 per day

Activity is above user's usual activity for the day

Figura 64. IRM File Uploads – Umbrales de eventos 5

Como sus nombres indican, estos umbrales detectan si el usuario ha involucrado en sus secuencias de acciones, actividad relacionada con archivos con contenido sensible, prioritario o dirigido a un dominio no permitido.

Lo descrito anteriormente corresponde a los eventos desencadenantes, los cuales únicamente indican al sistema cuándo debe iniciarse el análisis del comportamiento de un usuario concreto. Sin embargo, el hecho de que se genere una alerta no depende directamente de estos eventos iniciales, sino de si el usuario supera los umbrales establecidos en

determinados indicadores de riesgo, que son los que realmente determinan si la actividad observada debe clasificarse como potencialmente peligrosa o anómala.

Los indicadores de riesgo que he seleccionado para las políticas IRM los he explicado previamente, los que he escogido en concreto para esta política son los siguientes:

- Using a browser to upload files to the web
- Entering risky prompt in Copilot
- Receiving sensitive response from Copilot

Indicators

The following indicators are used to generate alerts for the activity detected by the policy template you selected.
[Learn more](#)

Total indicators selected: 5/90


 Unable to select some indicators? This is because they're currently turned off in your organization. To make them available to select, you can turn them on now.

[Choose indicators](#)

Office indicators (0/29 selected) 

Device indicators (1/15 selected) 

- Select all
- Printing files
- Creating or copying files to USB
- Mounting USB to a device
- File copied to remote desktop session
- Creating or transferring files to a network share
- Using a browser to upload files to the web

Microsoft Copilot experiences (preview) (2/2 selected) New 



- Select all
- Receiving sensitive response from Copilot 
- Entering risky prompt in Copilot 

Figura 65. IRM File Uploads – Indicadores de riesgo

Estos indicadores son los que realmente definen cuándo se va a generar una alerta y de qué nivel será según la cantidad de eventos generados y el umbral definido para cada indicador. En mi caso, he definido el mismo umbral para todos los indicadores y se generarán alertas de File Uploads a partir de los siguientes umbrales de cualquier indicador de riesgo:

- Alertas Low: >1 Alerta/día
- Alertas Medium: >3 Alertas/día
- Alertas High: >5 Alertas/día

Entonces, voy a colocar un ejemplo práctico para comprender cuándo se generará una alerta File Uploads:

Un usuario en su jornada laboral abrió tres documentos de la empresa almacenados en su OneDrive. A ese documento le quitó la etiqueta de confidencialidad, la subió a ilovepdf.com y luego la eliminó del OneDrive. Esta secuencia de acciones indicaría al sistema que debe comenzar a investigar al usuario ya que supera el umbral de eventos desencadenantes.

Luego de esto, el usuario envía cinco documentos a un dominio cualquiera, por ejemplo, web.whatsapp.com. Esto estaría superando el umbral de >3 eventos diarios del indicador de riesgo “Usar un navegador para subir archivos a la web” y se generaría por lo tanto una alerta de nivel High.

10.2.4.3.2 External emails with attachments

La finalidad de esta política es detectar la actividad de un usuario relacionada a envíos de emails con documentos adjuntos a dominios externos.

Un aspecto clave a resaltar es que, si bien la política de Data Loss Prevention realiza acciones al enviar correos a dominios externos, tienen diferentes funcionalidades, ya que la política DLP no realiza un conteo de las alertas generadas, IRM sí y además según el número de alertas generadas y su contenido, las clasifica como riesgo bajo, medio o alto. La implementación de la DLP fue realizada para llevar una acción de estrategia educativa a los trabajadores, la creación de esta alerta IRM se realiza por la capacidad de correlación entre múltiples acciones a lo largo del tiempo y detectar así patrones de comportamiento posiblemente peligrosos.

El evento desencadenante que la política debe detectar para comenzar a evaluar la actividad del usuario es “Sending email with attachments to recipients outside the organization”:

Choose triggering event for this policy

Choose one or more triggering events to determine when a policy will begin assigning risk scores to a user's activity. [Learn more](#)

User matches a data loss prevention (DLP) policy
Policy will start assigning risk scores when a user performs an activity matching the DLP policy you select. The DLP policy must be configured to generate 'High' severity incident reports. [Learn more about DLP policy requirements.](#)

Select a DLP policy

User performs an exfiltration activity
Policy will start assigning risk scores when specific thresholds are detected for activity relating to the following indicators:

Select which activities will trigger this policy

Downloading content from SharePoint

Sending email with attachments to recipients outside the organization

Printing files

Creating or copying files to USB

Using a browser to upload files to the web

Figura 66. IRM External emails with attachments – Eventos desencadenantes

El número mínimo de actividades que debe detectar para comenzar la monitorización son seis en total al día, dos conteniendo información sensible, dos con contenido prioritario y dos al día a dominios no permitidos.

Choose thresholds for triggering events

The policy will start assigning risk scores to activity only when specific thresholds are met for the exfiltration activities you selected as the triggering event. Thresholds are based on the number of events recorded for an activity per day. You can use recommended thresholds or specify your own.

Apply built-in thresholds RECOMMENDED
 Choose your own thresholds

Sending email with attachments to recipients outside the organization

Total number of activities
 per day

Number of activities for emails containing sensitive info types
 per day

Number of activities for emails matching priority content
 per day

Number of activities performed in which target is unallowed domain
 per day

Activity is above user's usual activity for the day

[Reset to defaults](#)

Figura 67. IRM External emails with attachments – Umbrales eventos desencadenantes

El indicador de riesgo que he seleccionado para esta política es “Sending email with attachments to recipients to outside the organization”

Sending Teams messages that contain sensitive info types
 Adding users outside the organization to a Teams private channel
 Adding users outside the organization to a Team
 Sharing file links with people outside organization in a Teams private channel
 Sharing folder links with people outside organization in a Teams private channel
 Sharing file links with people outside organization in a Teams chat
 Sending email with attachments to recipients outside the organization **(1/2 selected)** ^

Base: Sending email with attachments to recipients outside the organization
 Variant: Sending email to external

Device indicators (0/15 selected)

Physical access indicators (0/1 selected)

Figura 68. IRM External emails with attachments – Indicadores de riesgo

Los umbrales para definir si una alerta tiene nivel de riesgo bajo/medio/alto son:

- Alerta Low: >1 evento/día
- Alerta Medium: >3 eventos/día

- Alerta High: >5 eventos/día

10.2.5 Datalifecycle Management

Como se ha explicado anteriormente en este documento, el motivo para desarrollar esta parte de la solución del trabajo de fin de grado se basó en la idea de que la empresa de abogados tenía su sistema de almacenamiento de documentos referente a registros de forma local, almacenados de una forma muy ineficiente y desordenada.

Por ello, primeramente, se implementaron tres scripts, el primero organiza los documentos, el segundo sube esa estructura organizada a sitios de SharePoint y el tercero asigna propiedades mapeables (Crawled Properties) a los atributos de cada sitio de SharePoint para posteriormente mapearlas a propiedades gestionables (Refinable String) y aplicar un Adaptive Scope para crear políticas de retención con los diferentes Scopes en Data Lifecycle Management.

10.2.5.1 Script extract_fields.py

Este script permite analizar archivos PDF sin estructurar ubicados en un directorio local y clasificarlos según el tipo de trámite legal y el cliente correspondiente.

El script recorre un directorio que contiene documentos PDF. Para cada archivo, se realiza una lectura de su contenido textual mediante la librería fitz y se aplica una serie de reglas para:

1. Identificar el tipo de documento. Un ejemplo de ello sería: “Patente de invención”, “Marca nacional”, etc.
2. Extraer el nombre del cliente en base a ciertos campos específicos definidos para cada tipo de documento.
3. Crear una estructura de carpetas organizada por país, trámite y cliente.
4. Copiar el documento original a su carpeta correspondiente dentro del árbol estructurado “Arbol_Sites”.

En cuanto a la estructura de procesamiento, realicé lo siguiente:

- Definí un diccionario “campos” que asocia cada tipo de documento con un código de zona, por ejemplo "Spain_Patents", y los campos clave que se deben buscar en el PDF.
- La función ‘buscar_tipo’ detecta si un PDF contiene términos clave que identifiquen su tipo.
- Una vez identificado el tipo, se ejecuta extraer_campo para localizar el valor (cliente) asociado al documento.
- Por último, ‘agregar_nodo’ lo que hace es agregar el archivo dentro de un directorio en base a los datos del cliente.

Como resultado de su ejecución, el script genera un árbol de carpetas ordenado por país y cliente, donde cada documento queda archivado en función de su contenido. Esta estructura la utilizo luego con los scripts de automatización de SharePoint para generar los Sites correspondientes y cargar los documentos en las bibliotecas correctas.

A modo ilustrativo, se muestra el directorio de pruebas desordenado inicial:

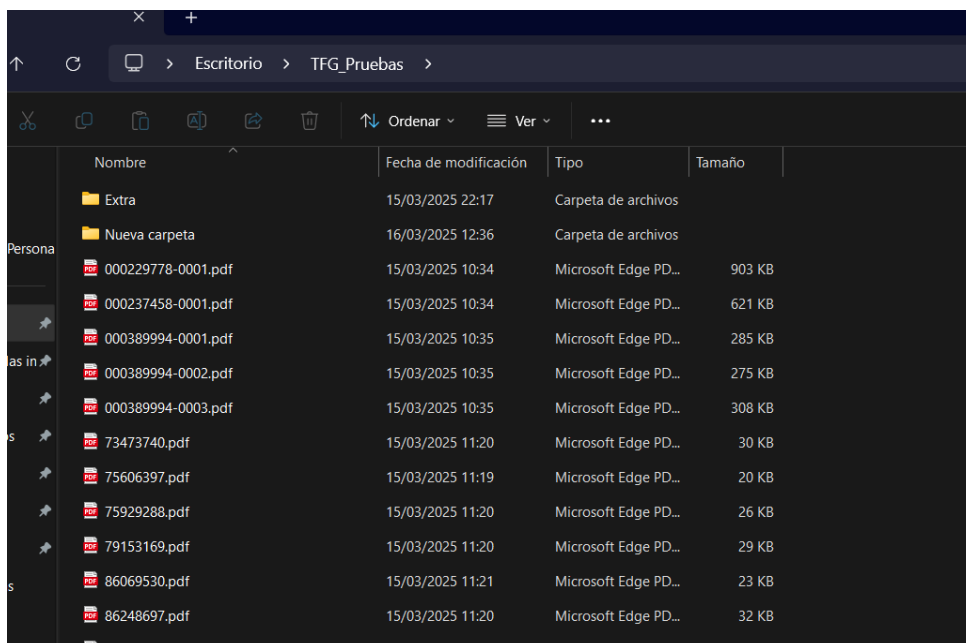


Figura 69. DLM – Almacenamiento local inicial

Tras aplicar el script, quedan organizados de la siguiente forma:

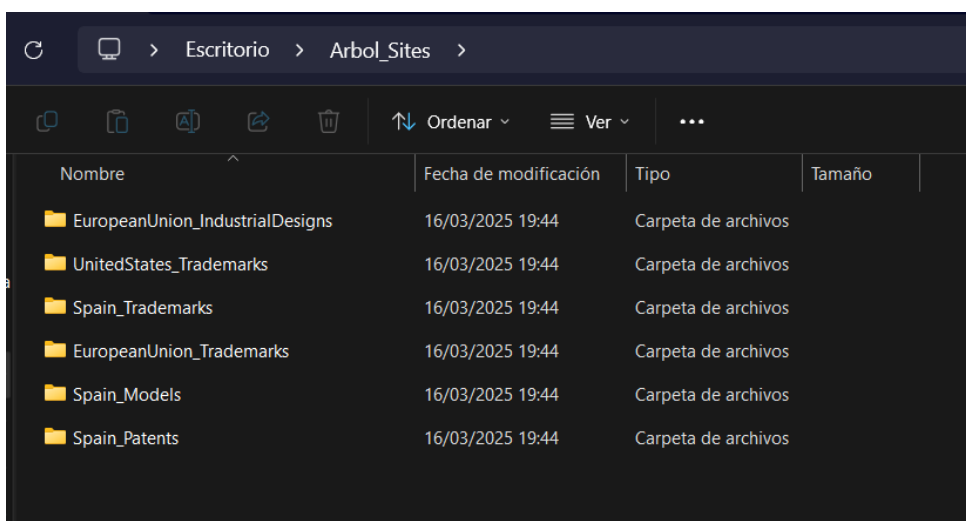


Figura 70. Ejecución script extract_fields, primer nivel

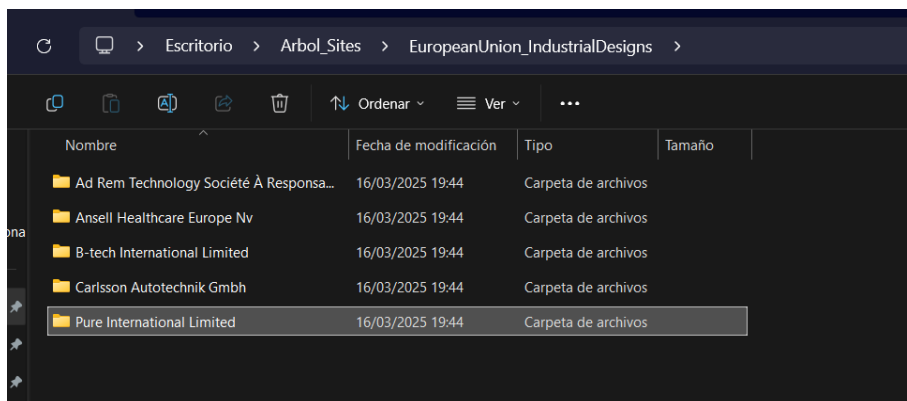


Figura 71. Ejecuci n script extract_fields, segundo nivel

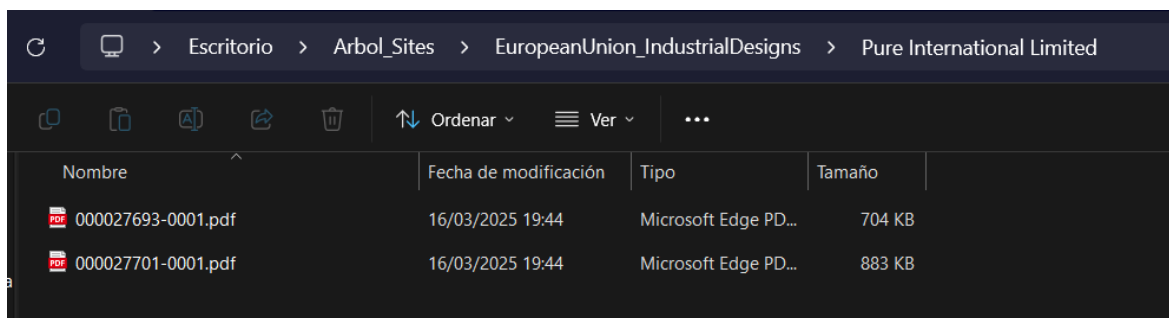


Figura 72. Ejecución script extract_fields, ficheros

10.2.5.2 Script create_sites.ps1

Para automatizar la creación de la arquitectura documental en SharePoint, se ha desarrollado el script create_sites.ps1, cuya finalidad es generar sitios (Sites), bibliotecas (Document Libraries) y cargar documentos al Site clasificados previamente en el sistema local.

La finalidad de este script es generar sitios (Sites) de SharePoint, bibliotecas y cargar todo el árbol de directorios que generé en el script anterior.

Este script está diseñado en PowerShell y se apoya en el módulo PnP.PowerShell, que es lo que nos proporciona poder interactuar con SharePoint Online y añadir Sites.

El script sigue las siguientes fases:

1. Función “verifyModule” comprueba si el módulo está instalado y en una versión igual o superior a la requerida (1.9.0). Si no es así, intenta instalarlo automáticamente o proporciona instrucciones al usuario para realizarlo manualmente.
2. AddPnPSites se encarga de verificar si existe un sitio con la URL esperada. En caso de que no exista, lo crea como Team Site asignando como propietario una cuenta administrativa. Esta función se ejecuta para cada carpeta padre en el directorio base “Arbol_Sites”, generando un sitio por país y trámite.
3. Ahora se crearía las bibliotecas necesarias para nuestra aplicación web. Estas son “Pending Appointments” y “Clients”. Es importante resaltar que estas bibliotecas se crean únicamente si no existen previamente, lo que permite la ejecución repetida del script sin causar duplicidades.
4. Finalmente tenemos la función UploadFilesToLibrary que lo que hace es recorrer los archivos dentro de cada carpeta del árbol de directorios “Árbol_Sites” y los sube a la biblioteca “Clients” del sitio que corresponda según la estructura de los directorios. La ruta relativa dentro del sitio respeta la estructura de carpetas original. Cada archivo es añadido con Add-PnPFile

Como resultado de la ejecución, el sistema genera automáticamente una estructura de sitios SharePoint organizada y alineada con el diseño previamente explicado en la sección de “Diseño” de este trabajo de fin de grado.

A continuación, se muestra el resultado esperado de la ejecución del script:

```

Creando/verificando sitio: EuropeanUnion_TradeMarks (https://m365x29060147.sharepoint.com/sites/EuropeanUnion_TradeMarks)
https://m365x29060147.sharepoint.com/sites/EuropeanUnion_TradeMarks
Sitio 'EuropeanUnion_TradeMarks' creado correctamente.
Creando la biblioteca 'Clients' en https://m365x29060147.sharepoint.com/sites/EuropeanUnion_TradeMarks
Nueva biblioteca 'Clients' creada!
Creando la biblioteca 'Pending Appointments' en https://m365x29060147.sharepoint.com/sites/EuropeanUnion_TradeMarks
Nueva biblioteca 'Pending Appointments' creada!
MicroTech.pdf File 443844 16/03/2025 18:56:42
Archivo subido: Alcide Basso\MicroTech.pdf
Kleindienst Daten... File 500505 16/03/2025 18:56:46
Archivo subido: Banctec Ecm Solutions Gmbh\Kleindienst Datentechnik.pdf
CDTCabletech.pdf File 354653 16/03/2025 18:56:51
Archivo subido: Cable Design Technologies Inc\CDTCabletech.pdf
Chiraltech.pdf File 393029 16/03/2025 18:56:56
Archivo subido: Daicel Corporation\Chiraltech.pdf
DoxTech.pdf File 361329 16/03/2025 18:57:01
Archivo subido: Doxtech Llc\DoxTech.pdf
Genentech.pdf File 364180 16/03/2025 18:57:06
Archivo subido: Genentech Inc\Genentech.pdf
Drytech.pdf File 474553 16/03/2025 18:57:10
Archivo subido: Jepson Power Group S.a\Drytech.pdf
Concordia.pdf File 352448 16/03/2025 18:57:14
Archivo subido: Landa Urbieta José Ignacio\Concordia.pdf
Millitech.pdf File 388385 16/03/2025 18:57:19
Archivo subido: Millitech Inc\Millitech.pdf
PeopleToPeople.pdf File 364005 16/03/2025 18:57:23
Archivo subido: Oki Electric Industry Co. Ltd\PeopleToPeople.pdf
TechnicallyElite... File 275094 16/03/2025 18:57:28
Archivo subido: Technically Elite Inc\TechnicallyElite.pdf
HighTechProgramm... File 399995 16/03/2025 18:57:33
Archivo subido: Viessmann Generations Group Gmbh & Co. Kg\HighTechProgramm.pdf
    
```

Figura 73. Ejecución script create_sites

Tras ejecutar el script, se crean los sitios de SharePoint según la estructura del árbol de directorios generado previamente:

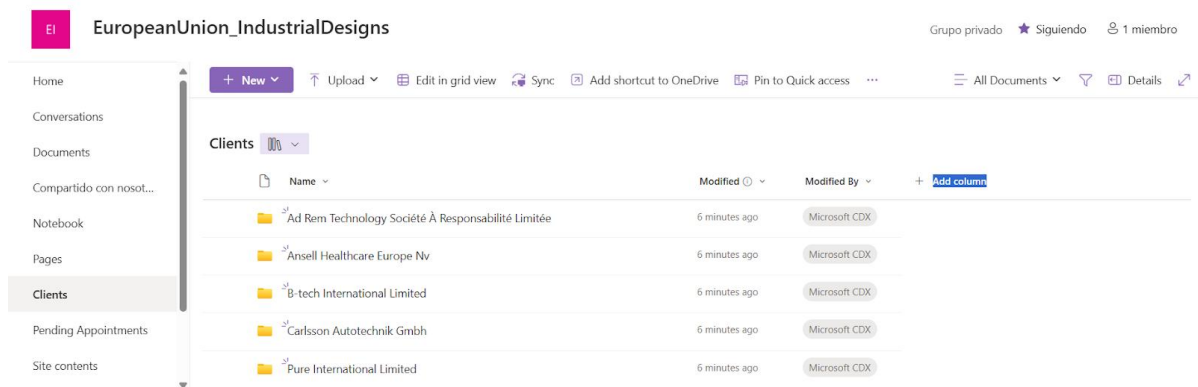


Figura 74. Ejecución script create_sites, primer nivel de directorios

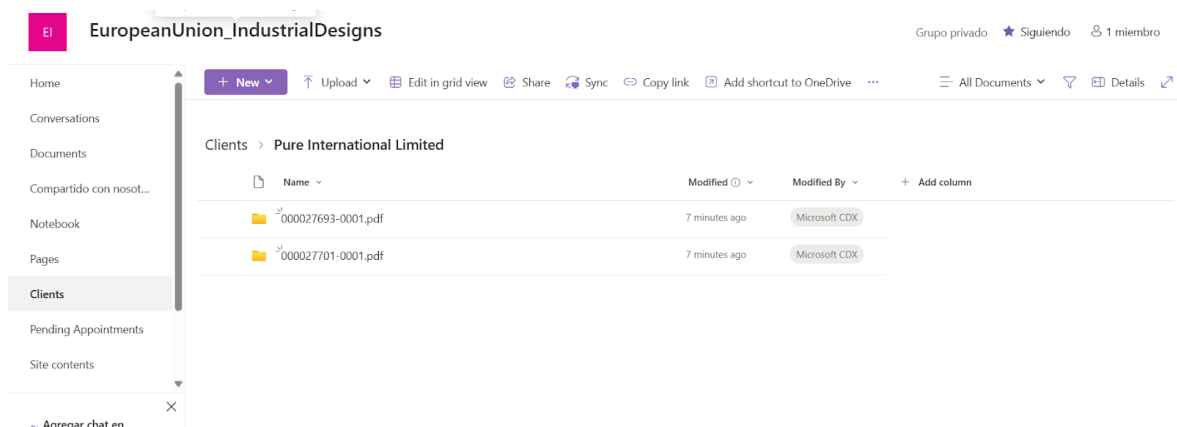


Figura 75. Ejecución script create_sites, segundo nivel de directorios

Y dentro de cada directorio, tendríamos todos los trámites realizados para ese registro en concreto. En este caso, solo un fichero pdf el cual es el registro realizado.

10.2.5.3 Script config_sites.ps1

Una vez creada la arquitectura de Sites en SharePoint y cargados los documentos correspondientes, es necesario aplicar un mecanismo de clasificación que permita vincular dichos Sites a políticas de retención mediante Adaptive Scopes en Microsoft Purview. Para ello, se ha desarrollado el script config_sites.ps1, cuya finalidad es asignar propiedades mapeables (crawled) personalizadas a cada Site según los criterios definidos previamente en el apartado de diseño.

El propósito del script es recorrer una lista de sitios proporcionada en un archivo .csv, verificar si ya tienen asignada una clave personalizada (por ejemplo, GeneralRetention), y, si no es así o si se permite sobrescribir, aplicar dicha propiedad con el valor correspondiente. Este proceso es fundamental para que las políticas de Data Lifecycle Management que van a crearse puedan filtrar y aplicar reglas sobre los Sites mediante los RefinableString Properties que van a configurarse.

El script realiza las siguientes tareas:

- La función verifyModule comprueba que el módulo PnP.PowerShell esté instalado y en su versión mínima requerida (1.9.0). Si no lo está, el script finaliza o sugiere al usuario actualizar el módulo.
- La función verifyCsvLocation comprueba que el archivo .csv esté disponible, correctamente formateado y genera un archivo de log (log_file.csv) donde se registran todas las acciones, incluyendo errores, sitios omitidos u operaciones exitosas.
- El script importa el archivo .csv, que debe contener al menos una columna con las URLs de los sitios (Url) y otra con la clave personalizada que se desea establecer. Luego recorre cada fila y:
 - Conecta al sitio correspondiente usando Connect-PnPOnline.
 - Verifica si la propiedad clave ya existe.
 - Si no existe, o si se permite sobrescribir (-overwrite), aplica la clave:valor mediante Set-PnPAdaptiveScopeProperty.
 - Si falla cualquier operación, lo registra en el archivo de log con una explicación detallada.

A lo largo de la ejecución se contabilizan los sitios completados, fallidos y omitidos. Al finalizar, el script muestra un resumen de la ejecución y destaca si hay errores a revisar en el log generado.

Una propiedad mapeable (crawled property) no es utilizable por sí misma, necesita mapearse con una propiedad gestionable (Refinable String). Una vez sea vinculada la clave:valor a la propiedad refinable por primera vez, no hará falta volver a hacer todo este proceso, con ejecutar el script ya será suficiente.

10.2.5.4 Proceso de vinculación de Crawled Property a Refinable String

A continuación, se enseñan los pasos a seguir para poder vincular una propiedad mapeable a una propiedad refinable:

1. Se debe ingresar al Search Schema y buscar la propiedad mapeable creada, en este caso fueron GeneralRetention y SpecialRetention:

Search

Managed Properties | **Crawled Properties** | Categories

Use this page to view or modify crawled properties, or to view crawled properties in a particular category. Changes to pr level.

Enter a name or partial name in the "Crawled properties" field to list matches. Click "Load All Crawled Properties" to view

The crawled property list displays each property's name and managed property mapping as "Property Name" and "Mapped to Property" was found. Crawled properties may have duplicate names, but they will belong to different categories. The last two columns property at a higher scope. If crawled properties with same name and property set Id exist in different scopes, the lower high is Site Collection -> Tenant -> System. An overridden property will be greyed out and can only be modified at the s

Filters

Crawled properties

Category

Show unaltered property names

PROPERTY NAME	MAPPED TO PROPERTY
GeneralRetention	
RetentionTime	
RetentionTime2	
SpecialRetention	

Figura 76. Propiedades mapeables

2. Se debe realizar la vinculación de la propiedad mapeable (crawled property) a la propiedad refinable (propiedad gestionable)

Search

Managed Properties | **Crawled Properties** | Categories

Use this page to view or modify the settings of this property. Note that the settings that you can adjust depend on your current authorization level.

Name and information

Name and description of the crawled property. This information on the crawled property is emitted by the filter or protocol handler.

Property Name: GeneralRetention

Category: SharePoint

Property Set ID: 00130329-0000-0130-c000-000000131346

Mappings to managed properties

Map this property to one or more managed properties.

RefinableString00(Text)

Figura 77. Vínculo entre propiedad mapeable y propiedad refinable

Una vez hecho esto, se debió esperar unas horas para que se hiciera todo el proceso interno de mapeo realizado por los indexadores de Microsoft.

A continuación, adjunto un resumen de las diferentes fases:

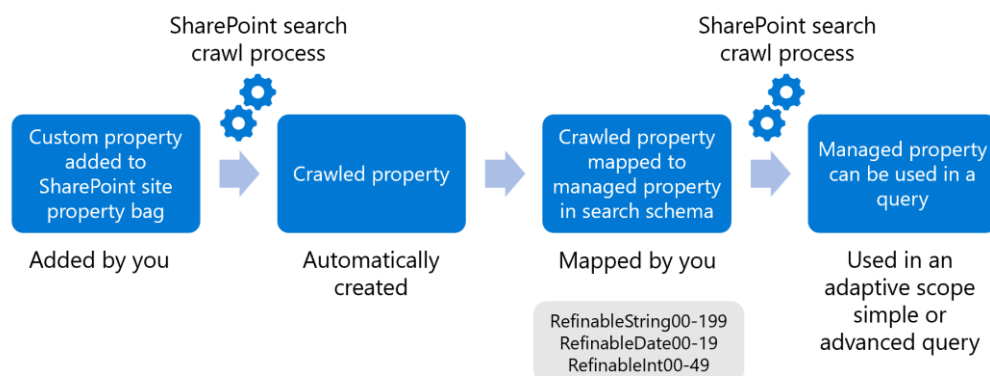


Figura 78. Resumen fases vínculo propiedad mapeable y propiedad refinable (Microsoft, 2022)

10.2.5.5 Creación de políticas de retención

Una política de retención tiene como objetivo definir el período durante el cual un contenido debe conservarse antes de ser eliminado de forma automática.

En el contexto de este proyecto, las políticas de retención se aplican a los documentos almacenados en bibliotecas de sitios de SharePoint, según el país de registro y el tipo de trámite legal correspondiente.

10.2.5.5.1 Creación de Adaptive Scopes

Para evitar aplicar manualmente cada política a sitios específicos, se ha optado por el uso de Adaptive Scopes (ámbitos adaptativos). Este enfoque permite definir reglas dinámicas que seleccionan automáticamente los sitios de SharePoint a los que debe aplicarse una política, en función de propiedades refinadas (managed properties) configuradas previamente.

En este caso, los Adaptive Scopes se basan en el valor de las propiedades RefinableString00, y RefinableString01, las cuales fueron asignadas a cada sitio tras la ejecución del script config_sites.ps1 y el mapeo correspondiente de su crawled property asignada. Dicha propiedad representa un par de clave:valor, donde la clave representa el tipo de retención a aplicar, y el valor representa la duración del periodo de retención. En este caso, fueron las siguientes:

- RefinableString00:3years
- ReginableString00:12years
- RefinableString01:1year

Es importante recordar que son dos Refinable Strings ya que el RefinableString00 está destinado a definir la duración general de un sitio según la duración de renovación del trámite y el RefinableString01 está destinado a definir un plus de un año más sobre la duración general si así se desea. Cada uno de estos pares clave:valor formará un ámbito adaptativo y requerirá la definición de una query de Microsoft (KQL) en cada uno de ellos.

Con propósitos de simplicidad, se adjunta solo un ejemplo de uno de los scopes definidos:

Create adaptive scope

Figura 79. Creación de query de Adaptive Scope

En resumen, se crearon tres scopes, uno para cada query respectivamente:

1. PatentModelDesign_Scope
2. Trademarks_Scope
3. PlusYear_Scope

10.2.5.5.2 Política de retención general de doce años

Esta política utiliza el scope de Trademarks_Scope. Lo único que tiene es la asignación del scope y la definición oficial del periodo de retención, el cual será doce años. Aplicará esta retención de forma dinámica a todos los sitios que tengan asignada la propiedad RefinableString00:12years.

Decide if you want to retain content, delete it, or both

Retain items for a specific period
Items will be retained for the period you choose.

Retain items for a specific period

of years months days

Start the retention period based on

At the end of the retention period

Delete items automatically

Do nothing

Retain items forever
Items will be retained forever, even if users delete them.

Only delete items when they reach a certain age
Items won't be retained, but when they reach the age you choose, we'll delete them from where they're stored.

Figura 80. Asignación temporal de doce años – DLM

10.2.5.5.3 Política de retención general de tres años

Esta política de retención usa el scope PatentModelDesign_Scope mediante su asignación en la configuración. Retiene por tres años el contenido de todos los sitios de SharePoint que tengan asignada la propiedad refinable 00 con el valor 3years:

Decide if you want to retain content, delete it, or both

Retain items for a specific period
Items will be retained for the period you choose.

Retain items for a specific period

of years months days

Start the retention period based on

At the end of the retention period

Delete items automatically

Do nothing

Retain items forever
Items will be retained forever, even if users delete them.

Only delete items when they reach a certain age
Items won't be retained, but when they reach the age you choose, we'll delete them from where they're stored.

Figura 81. Asignación temporal de tres años – DLM

10.2.5.5.4 Política de retención extra de un año

Esta política de retención es un poco diferente, ya que para dar flexibilidad al sistema se ha implementado de tal forma que sea una etiqueta. De esta manera, todos los sitios de SharePoint tienen su retención general aplicada dependiendo del sitio, pero, además, tienen la opción de usar esta etiqueta de retención en algunos documentos, lo que les permite almacenar por un año más todo lo que etiqueten.

Para ello usa el scope PlusYear_Scope, y define un periodo de retención de un año:

Define the retention period

Specify how long the retention period should be.

Retain items for

of years months days

Start the retention period based on

[+ Create new event type](#)

Figura 82. Asignación temporal de un año – DLM

Para mayor claridad, en la siguiente figura se enseña un ejemplo de cómo el usuario puede seleccionar una etiqueta para su documento, y entre las opciones está la etiqueta de retención Plus1 Year, la cual retiene los documentos por un año.

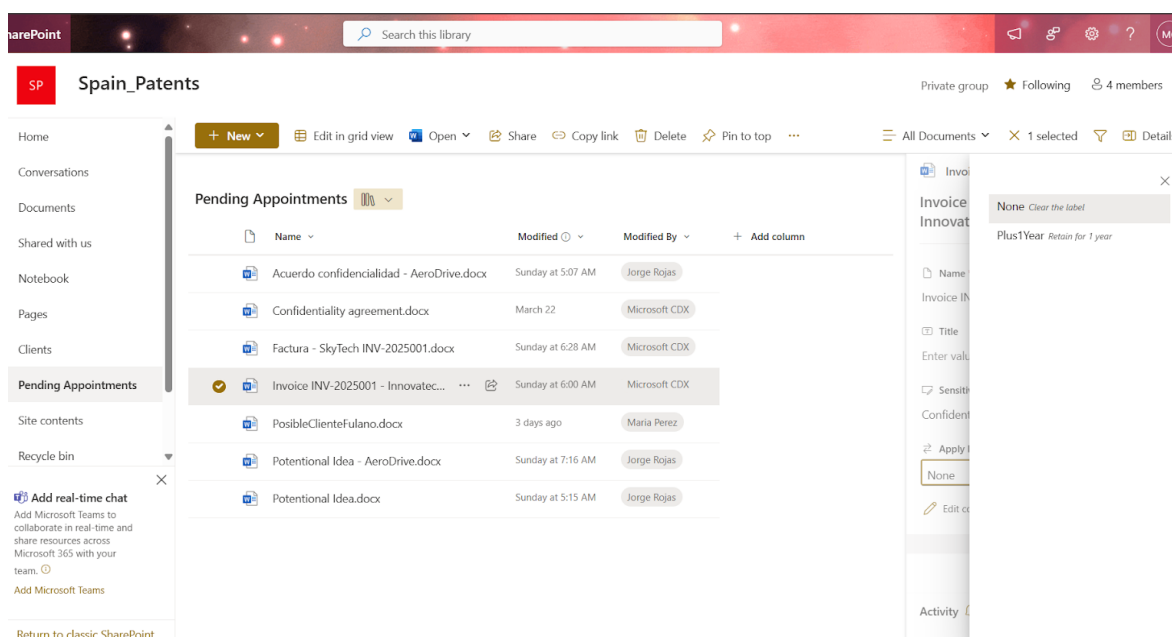


Figura 83. Etiqueta de retención – DLM

10.3 Fase de aplicación web

La aplicación web desarrollada ofrece la interfaz necesaria para gestionar la reserva de citas legales, la generación de acuerdos de confidencialidad y facturas, la protección de dicha información y su integración con servicios en la nube como Microsoft 365 para agendar citas en los calendarios y enviar correos de confirmación y subir los archivos generados a los espacios de trabajo compartidos SharePoint, además de integrarse con toda la protección de datos ofrecida por la fase de protección de datos anterior.

10.3.1 Backend

Esta fase se ha desarrollado íntegramente en Java con Spring Boot, estructurada en capas con separación de responsabilidades: controladores, servicios, DTOs, respuestas y seguridad, integrando Microsoft Graph API como pieza fundamental en el desarrollo de esta API REST.

10.3.1.1 Configuración del cliente de Microsoft Graph

El backend desarrollado para esta aplicación web requiere interactuar con diversos servicios de Microsoft 365, como el calendario de Outlook, el correo electrónico y sitios de SharePoint. Para ello, se ha utilizado el SDK oficial de Microsoft Graph para Java, el cual necesita de una configuración previa para operar correctamente de forma autenticada.

- **Registro de aplicación en Microsoft Entra ID**

Antes de poder consumir la API de Microsoft Graph, ha sido necesario registrar una aplicación en el portal de Microsoft Entra ID. Este registro permite que el backend actúe en nombre del usuario autenticado para realizar operaciones como crear eventos, enviar correos o subir archivos.

La aplicación fue configurada con los siguientes parámetros clave:

- **Tipo de permisos:** Se han asignado permisos delegados, lo que significa que la aplicación actúa en nombre del usuario que ha iniciado sesión.
- **Permisos concedidos:** Calendars.ReadWrite, Mail.Send, Sites.ReadWrite.All

- **Método de autenticación:** Se ha configurado el cliente de Microsoft Graph mediante el flujo de credenciales de cliente, en el que la aplicación se autentica directamente ante Microsoft Entra ID usando su `clientId`, `clientSecret` y `tenantId`. Esto permite realizar operaciones sin intervención del usuario.

10.3.1.1.1 Clase *GraphClientConfiguration.java*

Esta clase encapsula la configuración de autenticación e inicialización del cliente de Graph:

- Define un `GraphServiceClient`, inyectado mediante Spring con `@Bean` y reutilizado en toda la aplicación.
- Utiliza un proveedor de credenciales (`TokenCredentialAuthProvider`) construido sobre un objeto `ClientSecretCredential`, que se configura con el Client ID, Client Secret y Tenant ID de la aplicación registrada en Microsoft Entra ID. Solicita los scopes necesarios definidos en el registro de la aplicación.

Esta arquitectura permite que todos los servicios del backend que necesitan interactuar con Microsoft 365 lo hagan de forma centralizada y segura, sin duplicar lógica de autenticación.

El cliente configurado en `GraphClientConfiguration` es inyectado y utilizado por los siguientes servicios:

1. **CalendarService:** Crea y consulta eventos en calendarios de empleados.
2. **EmailService:** Envía correos electrónicos de confirmación a clientes y abogados.
3. **SharepointUploadService:** Realiza la subida de documentos a bibliotecas de SharePoint y gestiona carpetas de clientes.

10.3.1.2 Controladores

Los controladores son las clases que reciben las peticiones HTTP del cliente frontend y delegan la lógica en los servicios correspondientes. Todos los controladores están definidos como `@RestController` y exponen rutas bajo `/api/v1`.

10.3.1.2.1 *CalendarController.java*

Este controlador permite lo siguiente:

- Consultar la disponibilidad de los empleados de un determinado departamento (GET `/calendar/{departmentId}`).
- Crear un evento en el calendario de un abogado (POST `/calendar`).

Este último requiere autenticación (rol PAID) y valida el cuerpo de la petición usando `NewEventDTO`, que está anotado con restricciones como `@NotNull` o `@NotBlank`.

10.3.1.2.2 *PaymentController.java*

Expone un único endpoint para simular el pago de la cita (POST `/payment/`).

A partir del nombre recibido en el cuerpo de la petición, se genera un token JWT firmado, que es devuelto en una cookie segura y HTTP-only. Este token identifica al cliente en los pasos posteriores del flujo, sin este token el cliente no podrá agendar visitas en el calendario del abogado, ni tampoco podrá actualizar SharePoint con un nuevo documento. Es una medida de control que garantiza que realiza una acción un usuario identificado por su nombre en este entorno simulado.

10.3.1.2.3 *SharepointUploadController.java*

Expone la ruta POST /sharepoint. Como se mencionó anteriormente, esta está protegida por autenticación PAID, y permite subir documentos al Site de SharePoint correspondiente. Recibe un UploadDocumentsDTO que contiene un DocumentDTO con los datos del proyecto y del autor, además de los datos del abogado con el que está reservando la cita y el país donde quiere realizar su registro.

10.3.1.3 Servicios

Los servicios contienen la lógica funcional de negocio. Cada uno de ellos encapsula una parte específica del flujo de trabajo.

10.3.1.3.1 *CalendarService.java*

Contiene dos métodos principales:

- `getAvailabilityByDepartment(String departmentId)`: consulta, mediante Microsoft Graph API, la disponibilidad horaria de todos los empleados de un departamento. El resultado se devuelve en un objeto `UserAvailability`, que representa las franjas ocupadas para cada empleado en un `TimeRangeResponse` junto con su nombre y email.
- `addEventToUserCalendar(NewEventDTO dto)`: crea un evento en el calendario de un empleado utilizando `GraphServiceClient`, configurando el cuerpo del evento y programando el bloque horario indicado. Tras haber conseguido agendar una cita en el calendario del abogado, al final de este método se llama al `emailService` para enviar un correo de confirmación a los usuarios involucrados.

10.3.1.3.2 *EmailService.java*

Este servicio se encarga de enviar un correo de confirmación de cita tanto al cliente como al abogado. Utiliza `GraphServiceClient` para construir el mensaje y definir a los destinatarios. Incluye como campos el nombre del autor, la fecha y el motivo del encuentro.

10.3.1.3.3 *PaymentService.java*

Simula el proceso de pago y genera un JWT con el nombre del cliente como identificador. Internamente utiliza `JwtUtils` para firmar el token y construir una cookie `HttpOnly`.

10.3.1.3.4 *DocumentGeneratorService.java*

Genera dos documentos en base a la información recibida y en base a las plantillas del sistema. Recoge las plantillas y rellena los espacios delimitados por `{{ $variable }}` y genera los siguientes documentos:

- Un acuerdo de confidencialidad.
- Una factura por el trámite solicitado.

Ambos documentos se devuelven como arrays de bytes, preparados para ser subidos a SharePoint con el `graphClient`. Se ha decidido que el formato a subir sea `.docx` ya que esto permite la integración con el proceso de clasificación y cifrado de una forma más sencilla.

10.3.1.3.5 *SharepointUploadService.java*

Este servicio cumple con muchas funcionalidades de la aplicación, se explican a continuación:

- `buildSiteName(String country, String service)`: Determina el sitio de SharePoint adecuado en base al país y trámite.
- Se conecta al Site de SharePoint usando `GraphClientConfiguration`.
- `uploadFiles`: Este método es el que se encarga de subir los documentos generados a la carpeta correcta dentro de la biblioteca “Pending Appointments”.

El sistema aprovecha las políticas de retención creadas y se aplican, así como también se etiquetan los documentos y en cuanto a los acuerdos de confidencialidad, restringe su acceso solo para los abogados pertenecientes al grupo `IntellectualPropertyAgent`.

10.3.1.4 DTOs (Data Transfer Objects)

Los DTOs definen los objetos que se reciben en los cuerpos de las peticiones. Todos ellos incorporan anotaciones de validación como `@NotBlank`, `@Email`, `@Size`, que permiten validar los datos antes de que lleguen al endpoint correspondiente.

10.3.1.4.1 *NewEventDTO.java*

Contiene:

- Correos del empleado y cliente.
- Nombre del autor.
- Fecha y hora de inicio y fin.
- Descripción de la reunión.

Se usa en la creación del evento de calendario.

10.3.1.4.2 *UploadDocumentsDTO.java*

Incluye:

- `DocumentDTO`.
- País del trámite.

Se utiliza para subir documentos al SharePoint.

10.3.1.4.3 *DocumentDTO.java*

Contiene la estructura de un documento de acuerdo de confidencialidad o factura. Estos campos se utilizan para posteriormente rellenar los espacios de los `.docx` plantilla y generar los documentos nuevos. Contiene campos como:

- Nombre del proyecto.
- Autor del proyecto.
- Tipo de documento.
- Descripción de la idea.

10.3.1.4.4 *PaymentDTO*

Contiene el campo:

- Nombre del cliente.

10.3.1.5 Clases de respuesta

Estas clases definen respuestas estructuradas.

10.3.1.5.1 UserAvailability.java

Agrupar la disponibilidad horaria de un empleado en un día determinado. Contiene:

- ID y nombre del empleado.
- Lista de TimeRangeResponse.

10.3.1.5.2 TimeRangeResponse.java

Define una franja horaria con:

- Start: fecha y hora de inicio (DateTimeResponse).
- End: fecha y hora de fin (DateTimeResponse).

10.3.1.5.3 DateTimeResponse.java

Encapsula campos date y time como cadenas.

10.3.1.6 Seguridad

El sistema de seguridad se basa en el uso de JSON Web Tokens (JWT) y cookies seguras. Toda petición a endpoints sensibles requiere la validación del token. A continuación, indico las clases relacionadas con la seguridad y enumero sus tareas:

10.3.1.6.1 JwtUtils.java

Genera y valida los JWT. Incluye:

- Firma con clave secreta.
- Extracción de nombre de usuario.
- Comprobación de expiración.

10.3.1.6.2 TokenFilter.java

Filtro que intercepta cada petición, lee el token desde la cookie y lo valida antes de permitir el acceso.

10.3.1.6.3 WebSecurityConfig.java

Define la configuración de Spring Security:

- Permite el acceso sin autenticación a ciertos endpoints.
- Requiere rol PAID para otros.
- Registra el filtro TokenFilter.

10.3.2 Frontend

El frontend de la solución ha sido desarrollado con React utilizando TypeScript, con una estructura modular que facilita la reutilización y el mantenimiento del código. Se han definido varios componentes principales en el directorio components y un conjunto de utilidades en api. A continuación, se explican los elementos más relevantes de la arquitectura.

10.3.2.1 Conexión con la API y manejo de datos

Para la comunicación entre el frontend y el backend se ha configurado un proxy en el archivo vite.config.ts que redirige las llamadas del cliente a los endpoints del backend Spring Boot. Esto permite evitar problemas de CORS durante el desarrollo.

Además, se ha centralizado la lógica de llamadas HTTP utilizando dos archivos:

- **axiosConfig.ts:** Define una instancia personalizada de Axios con una URL base (/api/v1) que permite abstraer las llamadas y aplicar configuraciones comunes como cabeceras o manejo de errores.
- **useFetchData.ts:** encapsula llamadas GET de manera asíncrona. Es utilizado, por ejemplo, en CalendarStep.tsx para obtener la disponibilidad horaria de los abogados.

10.3.2.2 Componentes principales

A continuación, se describen los componentes desarrollados y su papel dentro del flujo de la reserva:

- **AppointmentFlow.tsx**

Componente principal que gestiona todo el flujo de reserva utilizando un sistema de pasos. Internamente controla el estado global de la cita mediante useState y renderiza los distintos pasos (ServiceCard, CalendarStep, ClientFormStep, ConfirmationStep) según el progreso del usuario.

- **ServiceCard.tsx**

Componente de selección visual del tipo de trámite legal. Permite al usuario elegir entre opciones como marcas, patentes o modelos de utilidad. Se destaca por su diseño interactivo que muestra el botón de acción al pasar el cursor.

- **CalendarStep.tsx**

Muestra las fechas y franjas horarias disponibles para cada abogado en base al tipo de servicio seleccionado. Este componente realiza una llamada a la API /calendar/{departamento_empleado} mediante useFetchData, procesa los datos, y genera dinámicamente los slots disponibles a partir de la lógica de disponibilidad.

- **ClientFormStep.tsx**

Formulario que permite introducir todos los datos del cliente y del proyecto. Implementa validación básica de campos obligatorios. Este paso recopila los datos requeridos para generar la factura y el acuerdo de confidencialidad para el usuario que está agendando la visita.

- **ConfirmationStep.tsx**

Muestra un resumen completo de la cita y los datos introducidos. Permite confirmar la cita. Una vez confirmado los datos, se llama al {handleConfirm} presente en el AppointmentFlow, lo cual desencadena las llamadas finales al backend, lo que inicia el proceso completo de reserva, pago, generación de documentos, subida a SharePoint y envío de correos con todos los useState generados en los pasos anteriores los cuales están preparados para su uso.

- **Stepper.tsx**

Encargado de mostrar visualmente el progreso del usuario en los distintos pasos del flujo de reserva. Es un componente reutilizable que permite resaltar el paso activo y los completados.

10.3.2.3 Tipos e interfaces

El proyecto cuenta con un directorio llamado `types`, donde se centraliza la definición de todas las interfaces TypeScript utilizadas a lo largo de la aplicación. Este enfoque modular y tipado permite mantener la coherencia entre componentes, facilita el mantenimiento y reduce errores en tiempo de desarrollo.

10.3.2.3.1 `types/index.ts`

Este archivo contiene las estructuras de datos clave del frontend, reflejando los objetos que se reciben o se envían al backend. También define los modelos utilizados internamente en los estados de los componentes.

Entre las interfaces más relevantes se encuentran:

1. **NewEventDTO**

Define la estructura que se utiliza para crear un nuevo evento en el calendario de Microsoft 365. Incluye información tanto del cliente como del abogado:

- Email y nombre del empleado
- Email y nombre del cliente
- Fecha
- Datos de reunión

2. **DocumentDTO**

Representa los datos necesarios para generar los documentos PDF del acuerdo de confidencialidad y la factura:

- Datos del autor
- Información del proyecto

3. **LegalServiceType**

Es un tipo enumerado restringido que representa los posibles servicios legales disponibles:

- `'marcas'`, `'patentes'`, `'modelos'`, `'industrial'`

4. **LegalService**

Objeto que describe un servicio legal mostrado en la interfaz:

- `id` (clave única basada en `LegalServiceType`)
- Nombre, descripción e ícono (para renderizado visual)

11 Evaluación

En esta sección se evaluó las tres fases de forma independiente: Fase de protección de datos, fase de usuarios híbridos y fase de aplicación web. En las dos primeras fases se ha incorporado múltiples capturas para un mejor entendimiento del funcionamiento de la tecnología implementada, con sus respectivos cuadros resumen para los casos que no han sido adjuntados como captura, con la finalidad de simplificar la extensión de este documento.

11.1 Fase usuarios híbridos

Esta sección abarca todas las verificaciones realizadas para confirmar que la configuración de la identidad híbrida ha sido realizada con éxito.

- Verificaciones de que el equipo “Nicolle” se añadió correctamente al dominio: El usuario debe poder iniciar sesión en su ordenador bajo el nombre de su dominio.

1. Creamos un nuevo usuario en el Administrador del Servidor:

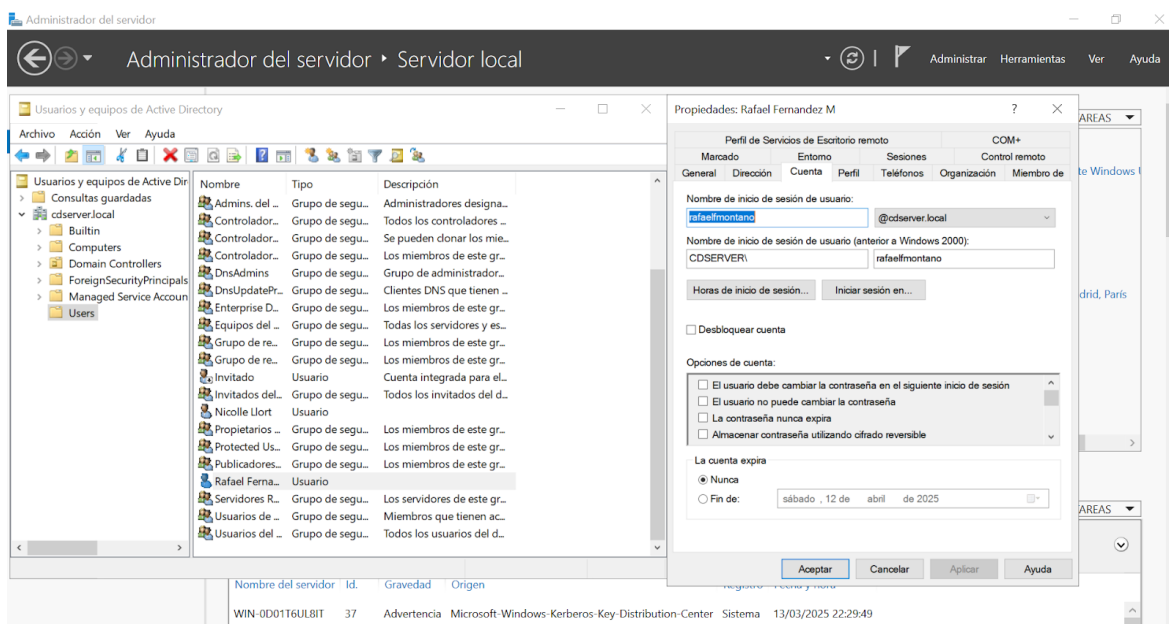


Figura 84. Administrador del servidor

2. Intentamos acceder con el usuario creado rafaelfmontano@cdserver.local a la máquina virtual “Nicolle” que se unió con anterioridad al dominio.
3. Se consiguió iniciar sesión y en la información del equipo de puede apreciar que el usuario es el mencionado anteriormente y que en el nombre completo del equipo figura el dominio de mi controlador de dominios:

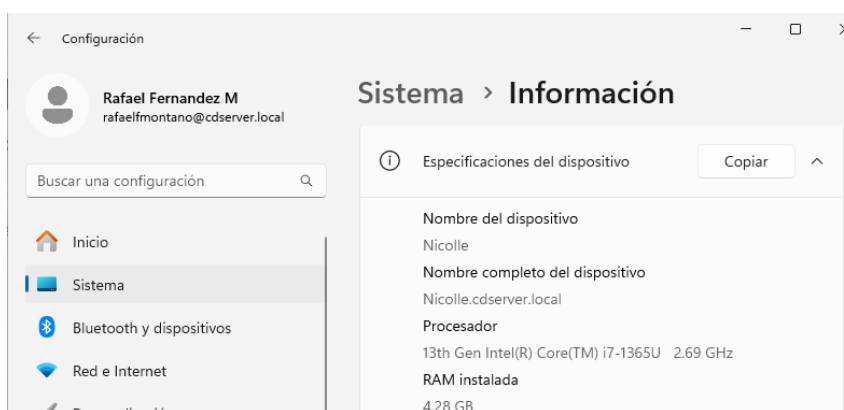


Figura 85. Información del sistema

- Verificar que el equipo “Nicolle” del empleado se encuentra añadido al dominio:
En el controlador de dominio, podemos ver en el apartado “Computers” de usuarios y equipos de Active Directory, se encuentra el equipo añadido correctamente:

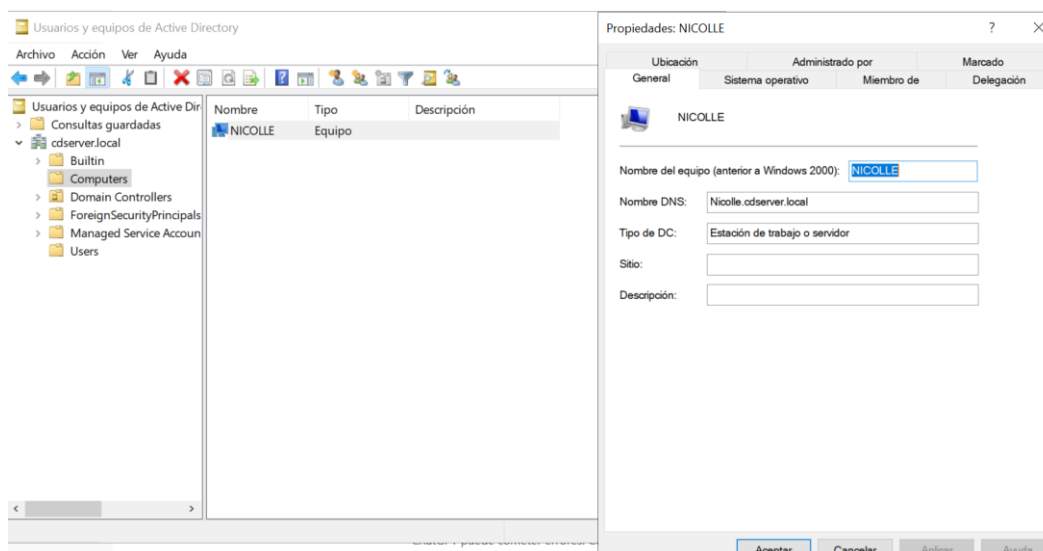


Figura 86. Equipo gestionado por el dominio

- Accesibilidad del cliente al servidor:
Se puede realizar ping al servidor desde el cliente:

```
C:\Users\rafaelfmontano>ping cdserver.local

Haciendo ping a cdserver.local [192.168.1.20] con 32 bytes de datos:
Respuesta desde 192.168.1.20: bytes=32 tiempo<1m TTL=128
Respuesta desde 192.168.1.20: bytes=32 tiempo=1ms TTL=128
Respuesta desde 192.168.1.20: bytes=32 tiempo=1ms TTL=128
Respuesta desde 192.168.1.20: bytes=32 tiempo=2ms TTL=128

Estadísticas de ping para 192.168.1.20:
    Paquetes: enviados = 4, recibidos = 4, perdidos = 0
              (0% perdidos),
    Tiempos aproximados de ida y vuelta en milisegundos:
        Mínimo = 0ms, Máximo = 2ms, Media = 1ms

C:\Users\rafaelfmontano>
```

Figura 87. Accesibilidad del cliente al servidor

- Correcta sincronización AD DS y tenant Entra ID:

El tenant de pruebas que se utilizó para realizar este trabajo de fin de grado incluía usuarios previamente creados. Podemos ver que estos no tienen la sincronización local habilitada porque solo son usuarios de originarios de cloud:




<input type="checkbox"/>	Nombre para mostrar ↑	Nombre principal de usu... ↑↓	Tipo de usuario	Sincronización local habilitada
<input type="checkbox"/>	 Adele Vance	AdeleV@M365x29060147...	 Miembro	No
<input type="checkbox"/>	 Alex Wilber	AlexW@M365x29060147...	 Miembro	No

Figura 88. Usuarios originarios de Cloud

Sin embargo, los usuarios que teníamos creados en el Active Directory local, tras hacer la identidad híbrida ahora aparecen también en Entra ID y además tienen activada la sincronización local por el entorno de identidad híbrida que hemos creado:



<input type="checkbox"/>	Nombre para mostrar ↑↓	Nombre principal de usu... ↑↓	Tipo de usuario	Sincronización local habilitada
<input type="checkbox"/>	 RF Rafael Fernandez M	rafaelfmontano@M365x2...	 Miembro	Sí

Figura 89. Usuarios sincronizados

- Habilitación SSO (Single Sign-On)

Además, si iniciamos sesión en la máquina local con el mismo usuario con el que iniciamos sesión en Azure (bajo el dominio del tenant @m365x29060147), obtenemos que es satisfactorio:

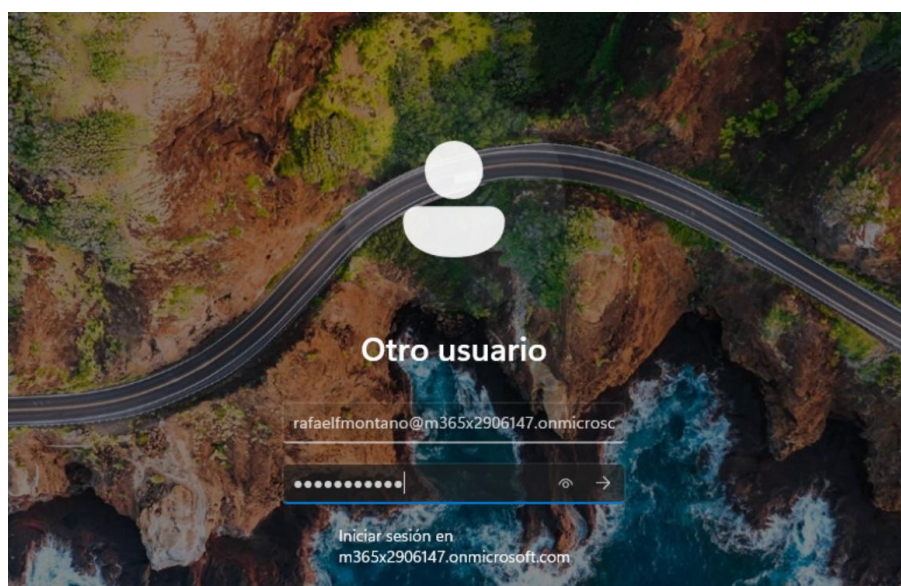


Figura 90. Single Sign-On

Se verifica que el SSO ha podido ser habilitado con éxito.

11.1.1 Juego de pruebas extra

A continuación, detallo pruebas que no han quedado cubiertas por las capturas de pantalla anteriormente explicadas, pero que también han sido importantes en la correcta verificación del sistema:

Prueba	Salida	Resultado
Ping desde el servidor al cliente	Se envían y reciben los paquetes	OK
Se inicia sesión con el usuario bajo el dominio del tenant en cualquier SaaS de la empresa	Se inicia sesión correctamente	OK
Se añade un nuevo usuario de forma local	Se sincroniza con Entra ID	OK
Se crea un nuevo grupo de forma local	Se sincroniza con Entra ID	OK
Se accede al navegador en todas las máquinas cliente antes de vincularlas al dominio	No puede realizar búsquedas ya que no tienen internet por sí solas	OK
Se accede al navegador en todas las máquinas cliente tras vincularlas al dominio	Pueden acceder a internet	OK

Tabla 1. Pruebas usuarios híbridos

11.2 Fase protección de datos

Esta sección abarca todas las pruebas realizadas para verificar el correcto funcionamiento de la fase de protección de datos.

11.2.1 Microsoft Information Protection

En esta sección se mostrarán las pruebas realizadas para las etiquetas Confidential non Encrypted, Confidential Encrypted, Intellectual Property Use e Internal Use implementadas. Para que estas pruebas funcionen, en la mayoría se requiere de Sensitive Info Types, con lo cual su funcionamiento queda verificado al ser probado el mecanismo de reconocimiento de patrones para aplicar las diferentes estrategias de las etiquetas.

11.2.1.1 Confidential non Encrypted y DLP Unencrypted Label

Para probar el correcto funcionamiento de esta etiqueta y política se ha etiquetado un documento con “Confidential Unencrypted”:

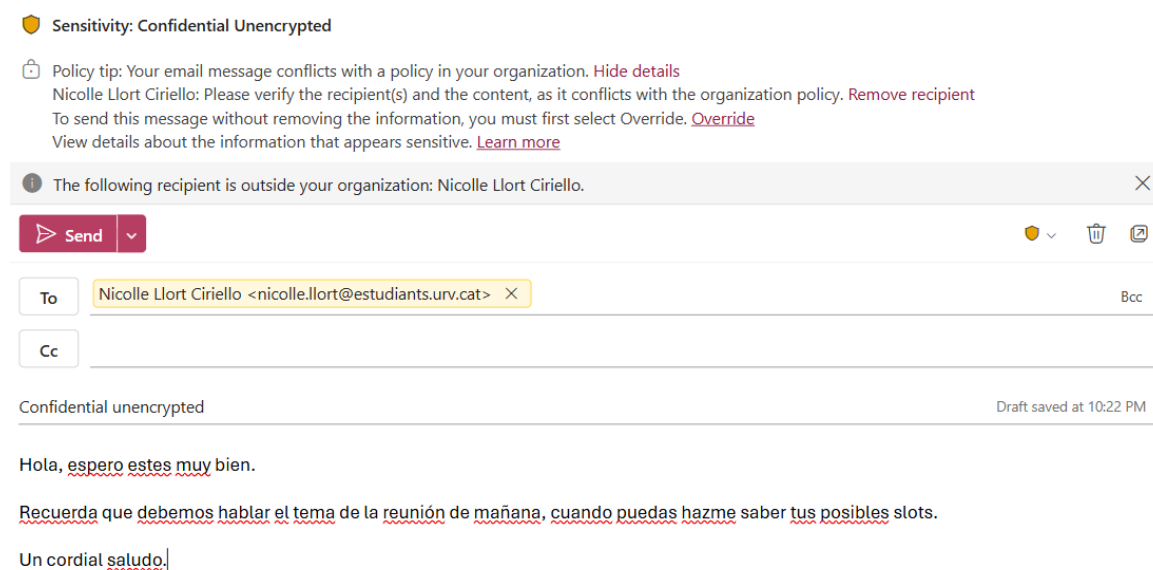



Figura 91. Ejemplo Confidential Unencrypted, policy tip

Como se puede apreciar, encima del correo ha aparecido un Policy tip que indica que se está entrando en conflicto con una política de la organización. Se explica que, para poder enviar este mensaje sin remover la información, deben hacer click sobre “Override”. Cuando hagan click sobre esta palabra, les aparecerá el siguiente recuadro:

Confirmar la invalidación

 **Sugerencia de directiva:** This item conflicts with a policy in your organization. [Mostrar detalles](#)

Su correo electrónico mensaje está en conflicto con una directiva de su organización. Has intentado enviar contenido clasificado como Confidential Unencrypted a uno o varios destinatarios externos, considere la posibilidad de quitar estos destinatarios o reemplazar el contenido.

Justificación

- Este destinatario tiene derecho a recibir este contenido
- Mi administrador ha aprobado el uso compartido de este contenido
- Otra justificación

Explique aquí su justificación empresarial.

Reconocimiento

- Confirmando que esta comunicación cumple con las directivas y estándares de mi organización.

Invalidar

Cancelar

Figura 92. Ejemplo Confidential Unencrypted, invalidación

Aquí deben indicar una de las opciones, y esto llegará al administrador de alertas encargado de analizar que esta información no represente un riesgo para la organización.

Una vez hecho esto, el correo se enviará sin problemas. Si el usuario decide ignorarlo e intenta enviar el mensaje sin justificar su acción, el envío será bloqueado y deberá volver a intentarlo, es por eso por lo que en este tipo de estrategias es importante las comunicaciones y píldoras informativas del equipo de gestión del cambio y tomarse su tiempo para que los empleados entiendan el proceso que deben atravesar y cómo hacerlo.

11.2.1.2 Confidential Encrypted

- Envío de correo con información sensible de DNI español:

Como se puede apreciar, aparece un aviso que recomienda cambiar la etiqueta por Confidential Encrypted:

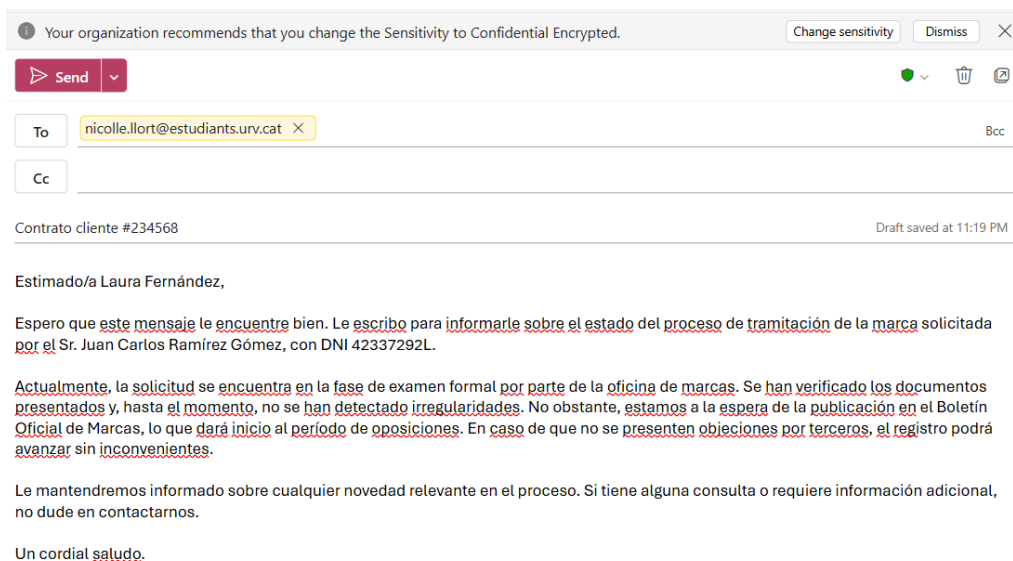


Figura 93. Recomendación cambio de etiqueta – Confidential Encrypted

Si lo aceptan haciendo click en “Change sensitivity” se aplica automáticamente:

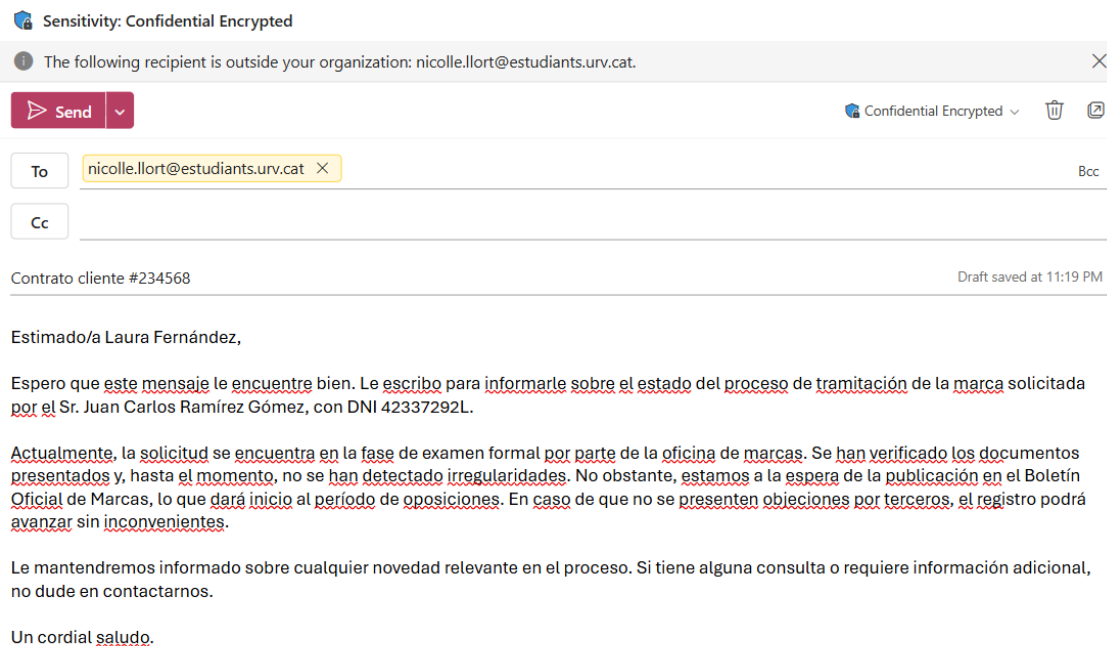


Figura 94. Ejemplo etiquetado - Confidential Encrypted

- Envío del correo anterior a un usuario externo a la organización:

Efectivamente, cuando el usuario externo intenta abrir el correo le aparece un mensaje de que el correo está bloqueado:

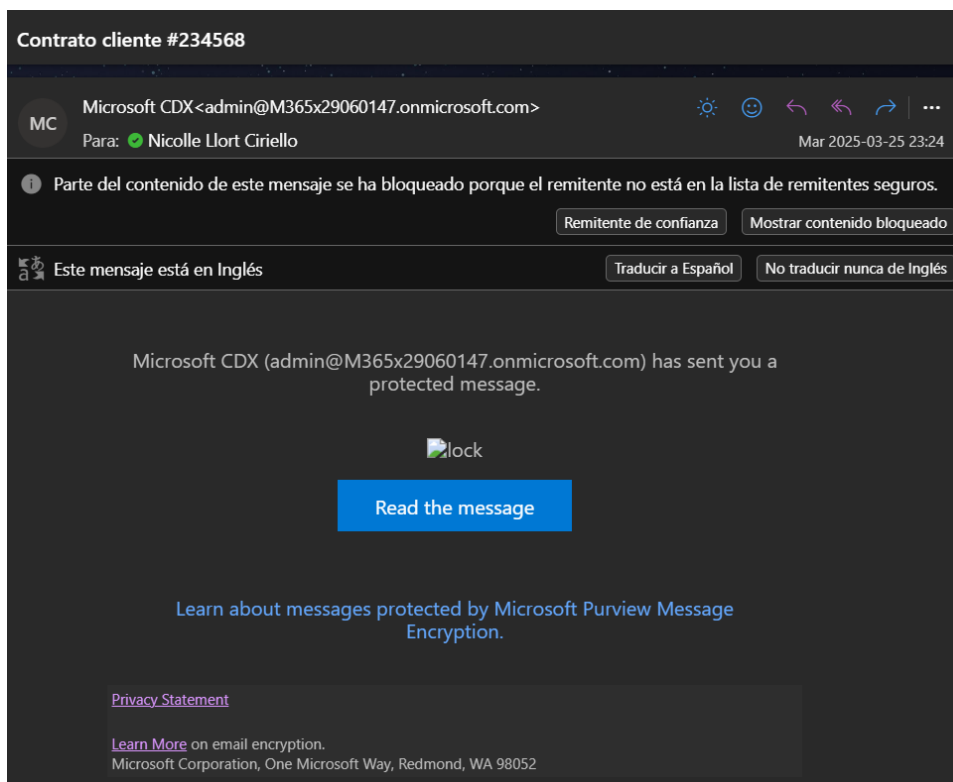


Figura 95. Acceso a correo bloqueado – Confidential Encrypted

Si el usuario hace click en “Read the message”, no puede abrir el mensaje:

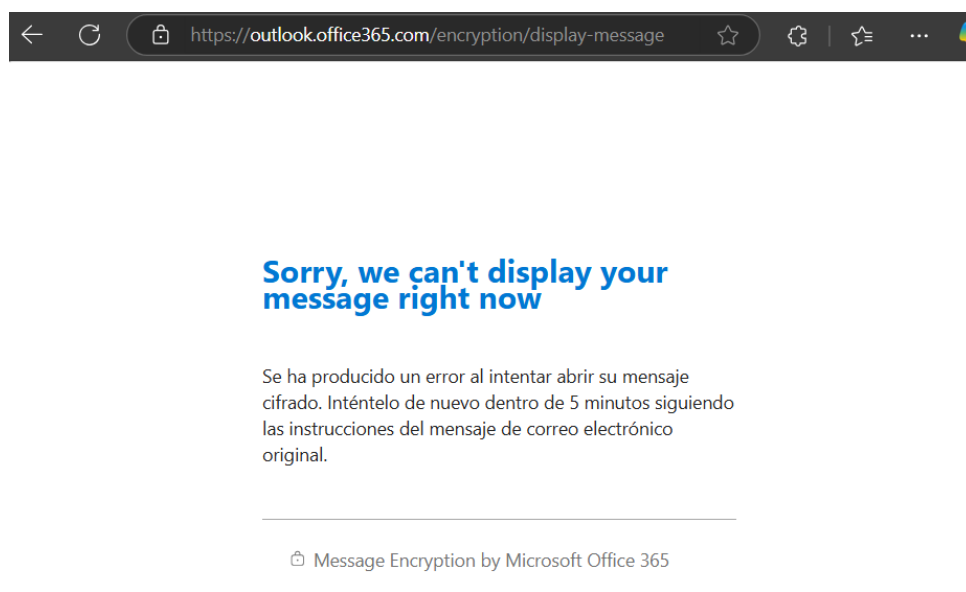


Figura 96. Acceso a correo bloqueado – Confidential Encrypted

- Prueba de que también recomienda en la redacción de documentos, pero en este caso el tipo de información sensible es la relativa a “Invoice”:

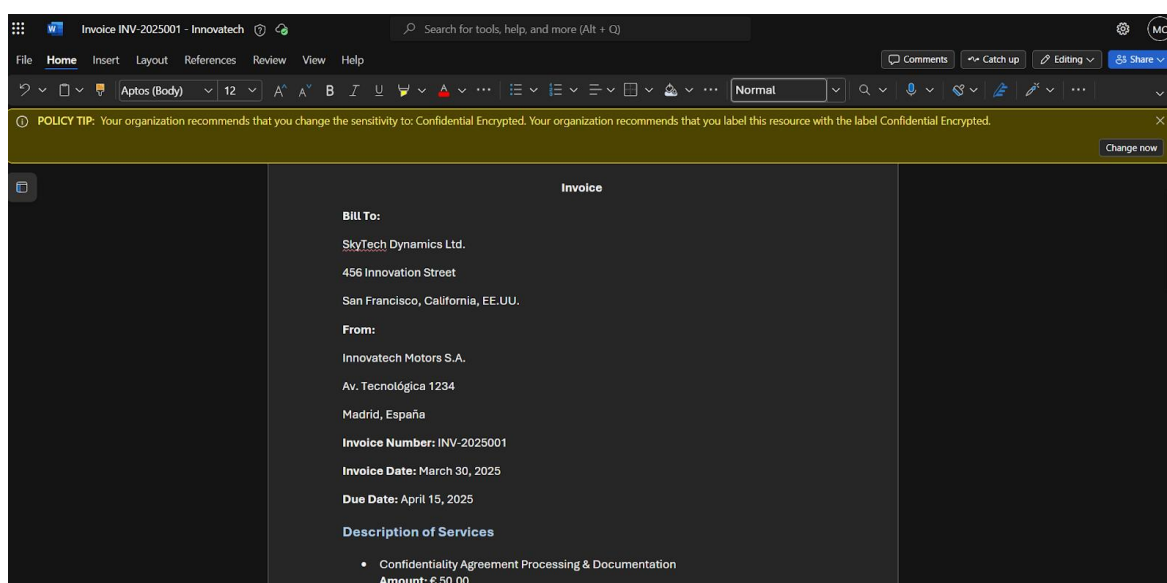


Figura 97. Acceso a correo bloqueado – Confidential Encrypted

Se probó con ambos patrones de divisa en el “Amount”, al igual que el mismo formato de documento, pero con su versión en inglés y con todos obtuvimos el resultado esperado.

- Se creó un documento de “Acuerdo de Confidencialidad” y se aplicó correctamente la etiqueta. En el Policy Tip se puede ver que se aplicó automáticamente:

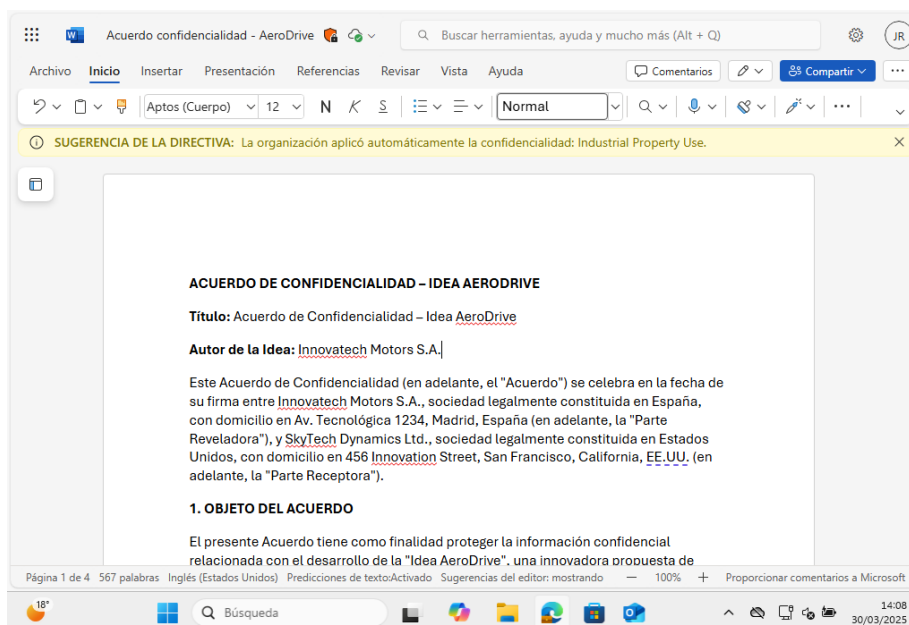


Figura 98. Aplicación a documento automática – Industrial Property Use

- Si una persona que no es miembro del grupo intenta abrir el documento, se deniega su acceso:

Sorry, you don't have permission to open this document

The document is protected by a rights management service, such as Azure Information Protection.

TECHNICAL DETAILS

GO BACK TO SITE

Figura 99. Negación de acceso – Industrial Property Use

- Se detecta y se aplica automáticamente la etiqueta si detecta el tipo de información sensible:

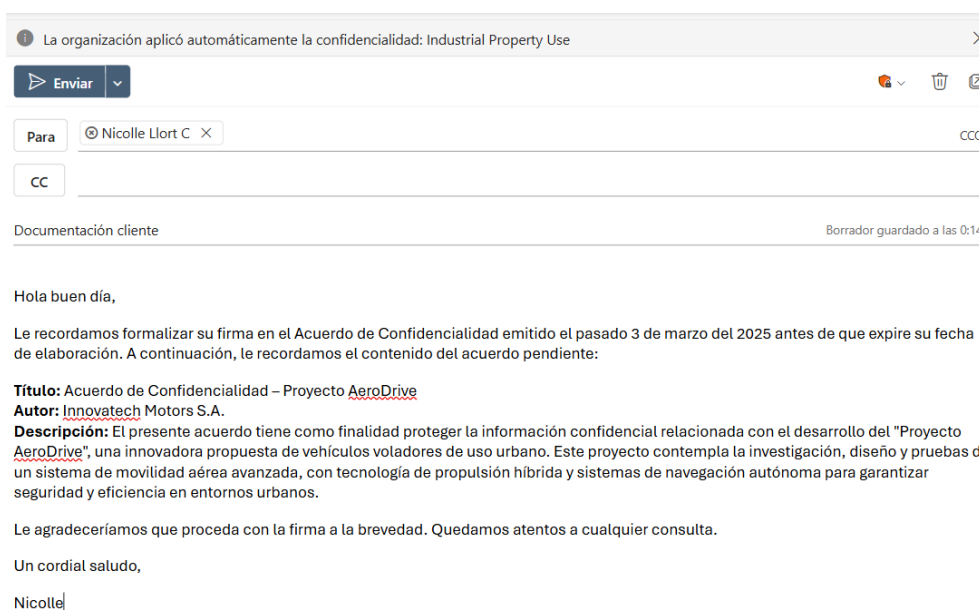


Figura 100. Aplicación a correo automática – Industrial Property Use

11.2.1.3 Industrial Property Use

Se ha probado que esta etiqueta sea la aplicada por defecto en todos los correos y documentos de la organización y que la información está correctamente cifrada.

11.2.1.4 Juegos de prueba extra

A continuación, se detallan los juegos de pruebas realizados que no están incluidos en las capturas anteriores:

Prueba	Salida	Resultado
Se escribió información sensible relacionada a "IBAN"	Se recomienda aplicar la etiqueta Confidential Encrypted	OK
Se escribió información sensible relacionada a "Número de pasaporte"	Se recomienda aplicar la etiqueta Confidential Encrypted	OK
Se escribió información sensible relacionada a "Tarjetas débito"	Se recomienda aplicar la etiqueta Confidential Encrypted	OK
Se escribió información sensible relacionada a "Invoice"	Se recomienda aplicar la etiqueta Confidential Encrypted	OK
Se reconoce contenido de acuerdos de confidenciales en inglés	Se aplica la etiqueta Industrial Property Use automáticamente.	OK

Tabla 2. Pruebas adicionales Microsoft Information Protection

11.2.2 Insider Risk Management

A continuación, se detalla de forma visual las pruebas realizadas para la implementación de la tecnología Insider Risk Management.

11.2.2.1 File uploads

Con propósitos de demostración, se han ejecutado dos grupos de prueba con dos usuarios diferentes, con la diferencia de que para la primera prueba se ha dejado la configuración de anonimización activada.

11.2.2.1.1 Alerta anónima

Se ha generado la siguiente alerta:

The screenshot shows a security alert interface. At the top, it says 'Alerts > File uploads'. The main title is '(38449312) File uploads'. There are buttons for 'Assign', 'Needs review', 'Confirm all alerts & create case', and 'Dismiss alert'. Below this, there's a summary: 'High severity Risk score: 75/100 Alert created on Mar 31, 2025 (UTC)'. The main section is 'Activity that generated this alert' with a sub-section 'Data exfiltration: Files uploaded to cloud storage' and a risk score of '75/100 High severity | Mar 30, 2025 (UTC)'. It lists 4 events: 'Files uploaded', 'Files uploaded to 1 unallowed domains', 'Files containing sensitive info, including: Invention_Proposal, IP', and 'Files that have labels applied, including: Industrial Property Use, Internal Use'. There are also 'Factors that impacted risk score' listed as 'Includes priority content (3 events)' and 'Includes unallowed domains (3 events)'. To the right, there are sections for 'Triggering event' (Using a browser to upload files to the web), 'User details' (Anonymized user ID), and 'User alert history' (1 alert for File uploads).

Figura 101. Vista previa anónima – File Uploads

Como se puede apreciar, el usuario es anónimo. Para generar esta alerta se han realizado las siguientes acciones:

- Para probar el indicador “**Entering risky prompt in Copilot**” se le ha introducido a Copilot un prompt en el que se le pide un resumen sobre un documento interno adjunto de la empresa el cual contiene información sensible:

The screenshot shows a Copilot chat interface. The user prompt is: 'Hola, dame un resumen de este documento: Proposal for Invention.docx'. The Copilot response is: '¡Claro! Aquí tienes un resumen del documento "Proposal for Invention":'. The response includes the title 'AeroDrive – Urban Air Mobility Solution', the author 'Innovatech Motors S.A.', and a description: 'AeroDrive es un concepto innovador destinado a revolucionar el transporte urbano mediante el uso de vehículos voladores. La idea principal es desarrollar un sistema de movilidad aérea avanzada que incorpore tecnología de propulsión híbrida y sistemas de navegación autónoma para garantizar tanto la seguridad como la eficiencia en entornos urbanos densamente poblados. Innovatech Motors S.A. pretende patentar esta idea para proteger su propiedad intelectual y comenzar los pasos necesarios hacia su desarrollo e implementación.' It also lists 'Características Clave de AeroDrive:'. There is a 'Ver indicaciones' link and a search bar at the bottom.

Figura 102. Introducción de prompt – File Uploads

Y luego otro prompt, en el que se le solicita que redacte un texto con información sensible (documentos de identidad, tarifas):

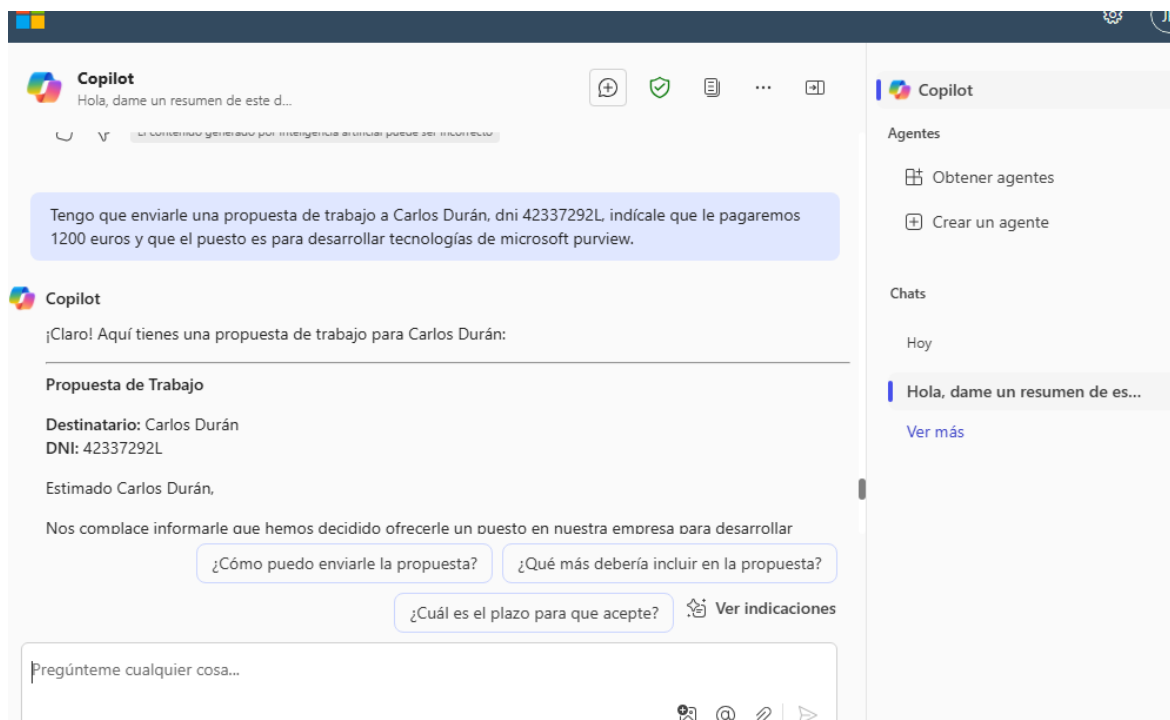


Figura 103. Solicitar información sensible – File Uploads

Con propósitos de que la alerta fuera activada, algunas pruebas se han ejecutado dos veces, si entramos al apartado “User Activity” de la alerta, vemos que esto ha generado los siguientes eventos relacionados a Risky prompt entered in Copilot de la alerta:

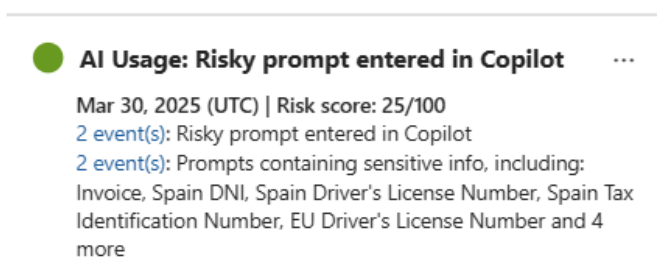


Figura 104. Resumen de eventos Copilot – File Uploads

Es importante señalar que Insider Risk Management te clasifica su detección de eventos primeramente de forma global (los dos primeros eventos) y luego te especifica más agrupaciones según la información que ha detectado dentro de esos mismos eventos, el cual sería el caso de la segunda agrupación de dos eventos que está listada en la imagen.

Si ingresamos al enlace de los dos eventos señalados:

Activity: Risky prompt entered in Copilot X		Date (UTC): 3/30/2025-3/30/2025 V		
Date (UTC)	Activity	Workload	Item type	File sensitivity label
<input type="checkbox"/> Mar 30, 2025 4:12 PM	Risky prompt entered in Copilot, Sen...	Copilot	Message	
<input type="checkbox"/> Mar 30, 2025 4:08 PM	Risky prompt entered in Copilot, Sen...	Copilot	Message	Industrial Property Use

Figura 105. Listado de alertas “Risky prompt entered” – File Uploads

Seleccionando el primer evento, podemos ver que es referente al ingreso de información relativa a nuestro tipo de información sensible “Invoice”:

Risky prompt entered in Copilot, Sensitive response received from Copilot

Activity details ^

Record id b96b7c29-1313-426a-8e2c-3e05107a149f	Date (UTC) Mar 30, 2025 4:12 PM
Workload Copilot	Operation CopilotInteraction
Activity Risky prompt entered in Copilot, Sensitive response received from Copilot	Application Office

About this record ^

Item type Message	Correlation ID 19:Bom7DP4r8KXApeJuoAyxd8ifVO KCbCMbQXGy5w7-o181@thread.v2
-----------------------------	--

Figura 106. Detalle de evento “Risky prompt entered” – File Uploads

Vemos que la alerta también es capaz de indicar qué tipo de información apareció en el prompt y si esta era información sensible:

About this record ^

Item type Message	Correlation ID 19:Bom7DP4r8KXApeJuoAyxd8ifVO KCbCMbQXGy5w7-o181@thread.v2
-----------------------------	--

Prompt details ^

Prompt Id
1743351122149

Sensitive info type

Name	Confidence	Count
UAE Passport Number	65	1
EU National Identification Number	76	1
EU Driver's License Number	66	1
Spain DNI	76	1
Spain Driver's License Number	66	1
Invoice	85	3
All Full Names	75	2

Response details ^

Response Id
1743351122596

Sensitive info type

Name	Confidence	Count
Invoice	85	3
All Full Names	75	2

Figura 107. Detalle de la información en el evento “Risky prompt entered” – File Uploads

Incluso te indica exactamente el prompt introducido:

Prompt

i Users with the Insider Risk Management Investigator role can view prompts and responses. [Learn more about Insider Risk Management roles](#)

Tengo que enviarle una propuesta de trabajo a Carlos Durán, dni 42337292L indícale que le pagaremos 1200 euros y que el puesto es para desarrollar tecnologías de microsoft purview.

Figura 108. Prompt de riesgo del evento “Risky prompt entered” – File Uploads

En el caso del segundo evento, es relacionado a la subida del documento con información sensible etiquetado con Industrial Property Use, te indica que el tipo de información detectada es “Invention Proposal”:

Risky prompt entered in Copilot, Sensitive response received from Copilot

Activity details ^

Record id 05cee75e-03b4-432d-a5b4- cadebad17ce7	Date (UTC) Mar 30, 2025 4:08 PM
Workload Copilot	Operation CopilotInteraction
Activity Risky prompt entered in Copilot, Sensitive response received from Copilot	Application Office

About this record ^

Item type Message	Correlation ID 19:zDsidnf_h3Zr5GcuHMBVDIVk5ue rBrYe3-NBnr-OOaQ1@thread.v2
-----------------------------	--

Prompt details ^

Prompt Id
1743350899305

Sensitive info type

Name	Confidence	Count
Invention_Proposal	85	3

Figura 109. Detección de Información sensible Invention Proposal– File Uploads

También indica el documento que fue compartido y si estaba etiquetado, con qué etiqueta:

About this record ^

Item type Message	Correlation ID 19:zDsidnf_h3Zr5GcuHMBVDIVk5ue rBrYe3-NBnr-OOaQ1@thread.v2
-----------------------------	--

Prompt details ^

Prompt Id
1743350899305

Sensitive info type

Name	Confidence	Count
Invention_Proposal	85	3

Response details ^

Response Id
1743350899659

Sensitivity label
Industrial Property Use

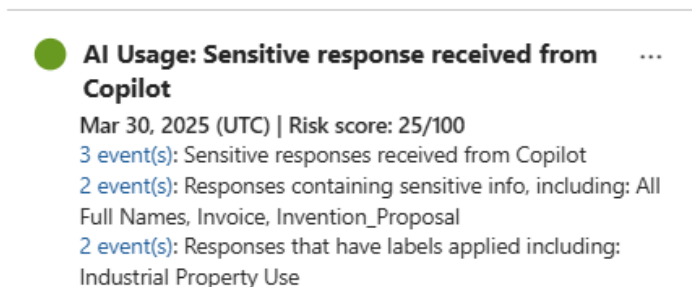
Accessed Resources ^

Accessed Resources

File name	File sensitivity label	Classifier matched	Sensitive info type (Name, Confidence, Count)
Proposal for Invention 1.docx	Industrial Property Use		

Figura 110. Detección de etiquetado sensible Industrial Property Use – File Uploads

- Para el indicador “**Receiving sensitive response from Copilot**” debido a que se le ha solicitado resúmenes/textos a Copilot referentes a la información que le he pasado, se han generado también respuestas sensibles por su parte, como se observó en las capturas anteriores, y esto es indicado también en los eventos que ha captado la alerta:



AI Usage: Sensitive response received from Copilot ...

Mar 30, 2025 (UTC) | Risk score: 25/100

3 event(s): Sensitive responses received from Copilot

2 event(s): Responses containing sensitive info, including: All Full Names, Invoice, Invention_Proposal

2 event(s): Responses that have labels applied including: Industrial Property Use

Figura 111. Evento Sensitive response Copilot – File Uploads

Si ingresamos a los tres eventos generados, nos encontramos con lo siguiente:

Activity: Sensitive response received from Copilot X Date (UTC): 3/30/2025-3/30/2025 v

Date (UTC)	Activity	Workload	Item type	File sensitivity label
<input type="checkbox"/> Mar 30, 2025 4:12 PM	Risky prompt entered in Copilot, Sen...	Copilot	Message	
<input type="checkbox"/> Mar 30, 2025 4:08 PM	Risky prompt entered in Copilot, Sen...	Copilot	Message	Industrial Property Use
<input type="checkbox"/> Mar 30, 2025 4:07 PM	Sensitive response received from Co...	Copilot	Message	Industrial Property Use

Figura 112. Listado de eventos Sensitive response Copilot – File Uploads

Seleccionando el primer evento, vemos que es referente al manejo de información sensible de tipo “Invoice” lo cual coincide con la solicitud que hizo el usuario de realizar una oferta de trabajo bajo datos sensibles:

Sensitive response received from Copilot

Activity details ^

Record id 8fd568c8-4d32-4022-b2e4-9f7c934bf6d1	Date (UTC) Mar 30, 2025 4:07 PM
Workload Copilot	Operation CopilotInteraction
Activity Sensitive response received from Copilot	Application Office

Figura 113. Detalle de actividad Sensitive response Copilot – File Uploads

Sensitive info type		
Name	Confidence	Count
UAE Passport Number	65	1
EU National Identification Number	76	1
EU Driver's License Number	66	1
Spain DNI	76	1
Spain Driver's License Number	66	1
Invoice	85	3
All Full Names	75	2

Figura 114. Detalle de información sensible Sensitive response Copilot – File Uploads

Y el segundo y tercer evento es relacionado a respuestas con información sensible referentes a información de tipo sensible “Industrial Property Use”:

Activity details

Record id 05cee75e-03b4-432d-a5b4-cadebad17ce7	Date (UTC) Mar 30, 2025 4:08 PM
Workload Copilot	Operation CopilotInteraction
Activity Risky prompt entered in Copilot, Sensitive response received from Copilot	Application Office

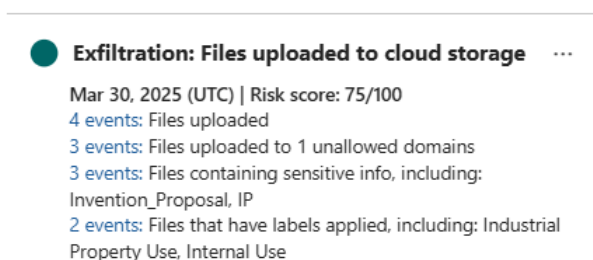
Figura 115. Detalle actividad sensible Industrial Property Use, Response Copilot – File Uploads

Prompt details ^			
Prompt id 1743350859863			
Response details ^			
Response Id 1743350860197			
Sensitivity label Industrial Property Use			
Sensitive info type			
Name	Confidence	Count	
Invention_Proposal	85	3	
Accessed Resources ^			
Accessed Resources			
File name	File sensitivity label	Classifier matched	Sensitive info type (Name, Confidence, Count)
Proposal for Invention 1.docx	Industrial Property Use		

Figura 116. Detalle del prompt, Sensitive response Copilot – File Uploads

- Para el indicador **“Using a browser to upload files to the web”** se realizó la subida de archivos a un dominio no permitido: web.whatsapp.com.

Si entramos al apartado “User Activity” podemos ver la actividad realizada por el usuario para generar la alerta:



Exfiltration: Files uploaded to cloud storage ...

Mar 30, 2025 (UTC) | Risk score: 75/100

4 events: Files uploaded

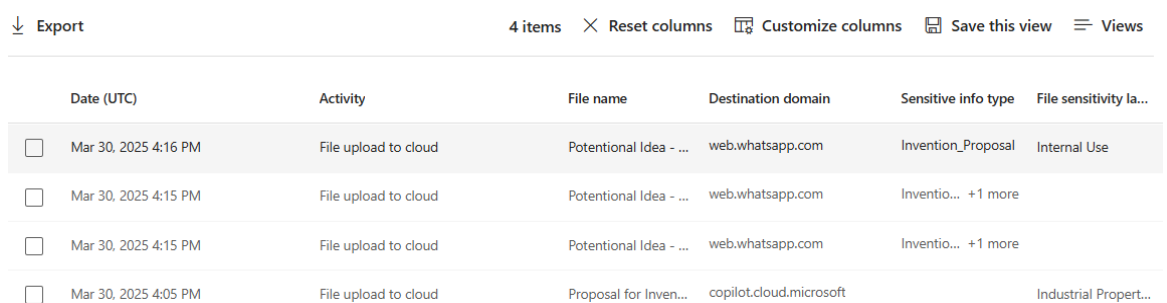
3 events: Files uploaded to 1 unallowed domains

3 events: Files containing sensitive info, including: Invention_Proposal, IP

2 events: Files that have labels applied, including: Industrial Property Use, Internal Use

Figura 117. Detalle de actividad usuario, subidas dominios web– File Uploads

Si ingresamos a los cuatro eventos generados:



Date (UTC)	Activity	File name	Destination domain	Sensitive info type	File sensitivity la...
<input type="checkbox"/> Mar 30, 2025 4:16 PM	File upload to cloud	Potential Idea - ...	web.whatsapp.com	Invention_Proposal	Internal Use
<input type="checkbox"/> Mar 30, 2025 4:15 PM	File upload to cloud	Potential Idea - ...	web.whatsapp.com	Inventio... +1 more	
<input type="checkbox"/> Mar 30, 2025 4:15 PM	File upload to cloud	Potential Idea - ...	web.whatsapp.com	Inventio... +1 more	
<input type="checkbox"/> Mar 30, 2025 4:05 PM	File upload to cloud	Proposal for Inven...	copilot.cloud.microsoft		Industrial Propert...

Figura 118. Listado eventos generados, subidas dominios web – File Uploads

Vemos que efectivamente, se han hecho tres subidas de archivos a web.whatsapp.com y una subida de archivos a copilot.cloud.microsoft la cual fue explicada anteriormente.

Las tres alertas se basan en la subida de archivos a whatsapp con información confidencial relacionada a “Invention Proposal”. Una de ellas es la siguiente:

File upload to cloud

Activity details ^

Record id 9c5bedf0-a8b3-46c1-8a30-3d3798628731	Date (UTC) Mar 30, 2025 4:16 PM
Workload Endpoint	Operation FileUploadedToCloud
Activity File upload to cloud	Application msedge.exe
Enforcement mode Audit	

Location details ^

Source location type Local	Originating domain m365x29060147.sharepoint.com
Destination location type Cloud	Destination domain web.whatsapp.com
Domain type	Client IP 83.40.132.247
Name web.whatsapp.com	Type Unallowed

Device details ^

Device full name
ada.cdserver.local

Figura 119. Detalle de la actividad File Upload to cloud – File Uploads

En la imagen se aprecian datos relevantes como el tipo de operación realizada, desde qué dispositivo se realizó y si el tipo de acceso era hacia un dominio no permitido.

Un detalle importante que considerar es que esta alerta es posible debido al despliegue explicado anteriormente del dispositivo ada.cdserver.local en el entorno de Purview.

Vemos también que sale reflejado el documento exfiltrado, el tipo de contenido en su interior y además se señala la ruta del documento que ha sido exfiltrado:

About this item ^

Item type File	Object ID C:\Users\jorgerojas\Downloads\Potential Idea - AeroDrive.docx						
File extension docx	File size 36 KB						
Sha1 db85f68517484859fa1c783f614a30d212c487f4	Sha256 0609871a1537d5176b60410826cb472a5d82708b380443a030d390e13aa144d9						
File name Potential Idea - AeroDrive.docx	Hidden false						
RMS encrypted false	Sensitive info type						
	<table border="0"> <thead> <tr> <th>Name</th> <th>Confidence</th> <th>Count</th> </tr> </thead> <tbody> <tr> <td>Invention_Proposal</td> <td>85</td> <td>3</td> </tr> </tbody> </table>	Name	Confidence	Count	Invention_Proposal	85	3
Name	Confidence	Count					
Invention_Proposal	85	3					

Figura 120. Detalle de los ítems File Upload to cloud – File Uploads

11.2.2.1.2 Alerta sin anonimizar

A propósito de demostración, se ha diseñado otra alerta donde sólo se intentó subir archivos a dominios no permitidos, ya que considero relevante poder monitorizar la realización de copias a dominios como google.drive y dominios no seguros como los conversores de texto a pdf que son tan utilizados hoy en día por los trabajadores administrativos sin considerar el riesgo que esto conlleva.

La alerta generada es la siguiente:

The screenshot shows a security alert interface. At the top, it says '(2024e1de) File uploads' with a search icon. There are buttons for 'Assign', 'Needs review', 'Confirm all alerts & create case', and 'Dismiss alert'. Below this, the alert is categorized as 'High severity' with a risk score of 75/100 and was created on Apr 2, 2025 (UTC). The main section is titled 'Data exfiltration: Files uploaded to cloud storage' with a severity of 75/100 High severity | Apr 1, 2025 (UTC). It lists 8 events: 1 file uploaded to 1 unallowed domain, 5 files containing sensitive info (including medical terms, physical addresses, diseases), and 5 files with labels applied (including industrial property use, public). Factors that impacted the risk score include priority content (2 events) and unallowed domains (1 event). On the right, there are sections for 'Triggering event' (Using a browser to upload files to the web), 'User details' (Nicolle Lloret C, Abogado, nicolle.M365x2906@147.onmicrosoft.com, Marcas), and 'User alert history' (File uploads, 1 alert).

Figura 121. Resumen de alerta sin anonimizar – File Uploads

Este es el detalle de los 8 eventos generados:

The detailed view shows an event titled 'Exfiltration: Files uploaded to cloud storage' with a risk score of 75/100. It lists 8 events: 1 file uploaded to 1 unallowed domain, 5 files containing sensitive info (including medical terms, physical addresses, diseases), and 5 files with labels applied (including industrial property use, public).

Figura 122. Detalle de eventos generados, alerta sin anonimizar – File Uploads

Vemos que efectivamente, ha sido detectada alertas en los dominios deseados:

Date (UTC)	Activity	File name	Destination domain	Sensitive info type	File sensitivity la...
Apr 1, 2025 10:16 PM	File upload to cloud	ES-2766923_B2.docx	drive.google.com		Industrial Propert...
Apr 1, 2025 10:15 PM	File upload to cloud	ES-2766923_B2.docx	www.ilovepdf.com		Industrial Propert...
Apr 1, 2025 10:14 PM	File upload to cloud	ES-2766923_B2.docx	online-file-converter.com		Industrial Propert...

Figura 123. Listado de eventos generados, alerta sin anonimizar – File Uploads

A modo de ejemplo, este es el detalle de la primera alerta, donde detalles importantes a considerar es que ya aparece el nombre del actor upn, aparece la ruta del fichero exfiltrado y el dominio en el que se ha hecho la exfiltración:

File upload to cloud Next item

Activity details

Record id 80a16443-c4b1-47d8-b70a-e351390c0d72	Date (UTC) Apr 1, 2025 10:16 PM
Workload Endpoint	Operation FileUploadedToCloud
Activity File upload to cloud	Application msedge.exe
Actor Upn nicolle@m365x29060147.onmicroso ft.com	Enforcement mode Audit

Location details

Source location type Local	Destination location type Cloud				
Destination domain drive.google.com	Domain type				
	<table border="0"> <thead> <tr> <th>Name</th> <th>Type</th> </tr> </thead> <tbody> <tr> <td>drive.google.com</td> <td>Unallowed</td> </tr> </tbody> </table>	Name	Type	drive.google.com	Unallowed
Name	Type				
drive.google.com	Unallowed				
Client IP 83.40.132.247					

Device details

Device full name
nicolle.cdserver.local

About this item

Item type File	Object ID C:\Users\nicolle\OneDrive - Contoso\ES-2766923_B2.docx
File extension docx	File size 506 KB
Sha1 1d75c1cb918c9bb9cc26aa5da69000 41535eab21	Sha256 5e6397b1fc7a6b06840f0527c4a75e 98fa34bf55c99240dbb4081d8e3eea b3e4
File name ES-2766923_B2.docx	Hidden false
RMS encrypted true	File sensitivity label Industrial Property Use

Figura 124. Detalle de actividad, alerta sin anonimizar – File Uploads

11.2.2.2 External emails with attachments

Para generar estas pruebas, se han enviado correos a dominios externos de alto riesgo como lo es mi dominio de gmail: nicollellort@gmail.com.

Se ha generado la siguiente alerta:

High severity Risk score: 75/100 Alert created on Apr 4, 2025 (UTC)

Activity that generated this alert [Reduce alerts for this activity](#)

Data exfiltration: Emails with attachments sent outside the organization
 75/100 High severity | Apr 2, 2025 (UTC)
 4 emails: sent to 1 recipient outside the organization
 4 emails: sent to 1 recipient from 1 unallowed domain
 4 emails: sent to free public domains
 3 emails: containing sensitive info, including: Invention_Proposal, Invoice
Factors that impacted risk score:
 Includes unallowed domains (4 events)

[View all activity](#)

Triggering event ⓘ
 Apr 3, 2025 (UTC)
 Sending email with attachments to recipients outside the organization

User details
 Maria Perez
 Abogado
 mariaperez@M365x29060147.onmicr...
 Marcas
[View all details](#)

User alert history
 Last 30 days
 External emails with attachme... 1 alert
[View full user history](#)

Figura 125. Detalle de la alerta External emails with attachments

La actividad que generó el usuario, en resumen:

Exfiltration: Emails with attachments sent outside the organization ...

Apr 2, 2025 (UTC) | Risk score: 75/100
 4 emails: sent to 1 recipient outside the organization
 4 emails: sent to 1 recipient from 1 unallowed domain
 4 emails: sent to free public domains
 3 emails: containing sensitive info, including: Invention_Proposal, Invoice

Figura 126. Detalle de actividad del usuario - External emails with attachments

Se puede destacar que de 4 eventos generados, todos han sido hacia un dominio no permitido y tres de estos correos han contenido información sensible, como Invention_Proposal o Invoice.

Los eventos generados fueron los siguientes:

Activity: Email sent to external recipient ✕		Date (UTC): 4/2/2025-4/2/2025 ▾	
Date (UTC)	Activity	Email subject	Email attach... Sensitive info type
<input type="checkbox"/> Apr 2, 2025 10:54 PM	Email sent to external recipient, f7bb...	P8	9
<input type="checkbox"/> Apr 2, 2025 10:50 PM	Email sent to external recipient, f7bb...	P6	9 Invoice +1 more
<input type="checkbox"/> Apr 2, 2025 10:32 PM	Email sent to external recipient, f7bb...	P4	9 Invoice +1 more
<input type="checkbox"/> Apr 2, 2025 10:30 PM	Email sent to external recipient, f7bb...	P3	9 Invoice +1 more

Figura 127. Listado de eventos de la alerta - External emails with attachments

Todas las alertas generadas fueron probadas bajo el mismo caso de uso, adjuntando 9 documentos con información sensible y siendo esto enviado a un dominio gmail. Una de estas alertas es la siguiente:

Email sent to external recipient, f7bb3e76-34c8-4876-820d-a7dbbab663ea

Activity details ^

Record id	Date (UTC)
820f20e4-068c-8f84-cdef-1b509a05f26b	Apr 2, 2025 10:30 PM
Workload	Operation
IrmHygiene	IrmHygieneData
Activity	Actor Upn
Email sent to external recipient, f7bb3e76-34c8-4876-820d-a7dbbab663ea	mariaperez@m365x29060147.onmicrosoft.com

Location details ^

Originating domain	Recipients				
m365x29060147.onmicrosoft.com	nicollellort@gmail.com				
Destination domain	Domain type				
gmail.com	<table border="0"> <tr> <td>Name</td> <td>Type</td> </tr> <tr> <td>gmail.com</td> <td>Unallowed</td> </tr> </table>	Name	Type	gmail.com	Unallowed
Name	Type				
gmail.com	Unallowed				

Figura 128. Detalles de la actividad - External emails with attachments

About this item ^

Item type	Object ID
Email	<PH0PR03MB659269C32BE3DEF776FAA039CAAF2@PH0PR03MB6592.namprd03.prod.outlook.com>
File name	Email network message ID
Factura - SkyTech INV-2025001.pdf,Potential Idea - AeroDrive.pdf,Potential Idea - AeroDrive.docx,Confidentiality agreement.docx,Factura - SkyTech INV-2025001.docx,Acuerdo confidencialidad - AeroDrive.docx,ES-2766923_B2(1).pdf,ES-2766923_B2.pdf,Acuerdo confidencialidad - AeroDrive.docx	2ee84b5b-84d4-4d9a-57f3-08dd7235f685
Email internet message ID	Email subject
<PH0PR03MB659269C32BE3DEF776FAA039CAAF2@PH0PR03MB6592.namprd03.prod.outlook.com>	P3

Figura 129. Detalles del ítem de la actividad- External emails with attachments

Email attachment count	Email attachment data		
9	File name	File extension	Sha256
			66bb32eea
			216eb8e8a
	Factura - SkyTech		183cfb264
	INV-2025001.pdf	pdf	7354b38fd
			6064a4b69
			b0d02bb7
			20055ad4689
			a0f08adf86
	Potential Idea - AeroDrive.pdf	pdf	c95753624
			c6f78feda1
			acd6a4d31
			c03f0d2cc
			ba9bac1c7
			0af08b29
			0609871a1
			537d5176b
	Potential Idea - AeroDrive.docx	docx	60410826c
			b472a5d82
			708b38044
			3a030d390
			e13aa144d9
			831f5e914
			8a7ed7919
	Confidentiality agreement.docx	docx	603d58a98
			6503681df
			4aae75c63
			1e79d7684
			3d659eb82c
			42290923e
	Factura - SkyTech		0e4fd1d7b
	INV-2025001.docx	docx	c0caf94563
			ac5f7fcf3b
			6d811d3de
			e8cd09708
			1c9319f2
			f9ebf729d
			5d91bfbc2

Figura 130. Resumen de los documentos adjuntos - External emails with attachments

En esta parte se puede apreciar el tipo de información que tenían los documentos:

Sensitive info type		
Name	Confidence	Count
Invoice	97	6
Invention_Proposal	85	9

Figura 131. Resumen de los documentos adjuntos - External emails with attachments

En las capturas hemos podido apreciar información valiosa que no podríamos saber si no hiciéramos uso de una herramienta como esta, hemos podido identificar exfiltraciones a correos externos, su contenido, el destinatario y el usuario que las ha provocado.

11.2.3 Datalifecycle Management

- Adición de propiedades mapeables a un Site de SharePoint tras ejecutar el script `config_sites.ps1`:

Se puede apreciar que existe la propiedad `GeneralRetention`, con el valor de tres años.

```
PS C:\Users\nicol\Desktop\TFG>> Connect-PnPOnline -Url "https://m365x29060147.sharepoint.com/sites/Spain_Models" -ClientId 09689b5d-4ed6-45ad-9592-7be4b83a2
2fe -Interactive
PS C:\Users\nicol\Desktop\TFG>> Get-PnPPropertyBag

Key                                     Value
---                                     -
GroupID                                f92f6f3d-05fc-42ce-9d7f-9e811b35635c
vti_sitemasterid                        8d0bf618-83d6-47ba-937f-e003feb345be
vti_categories                          Travel Expense\ Report Business Competition Goals/Objectives Ideas Miscellaneous Waiting VIP In\ Process Planning Schedule
vti_createdassociategroups              3;4;5
vti_defaultlanguage                      en-us
HomepageProvisioned                     1
contenttypesynctimestampversion         1
vti_approvallevels                      Approved Rejected Pending\ Review
taxonomyhiddenlist                      431ff79b-3e59-449b-90f4-b9da2250bdd8
UseFastCloneFromSiteMaster              TRUE
vti_associategroups                      5;4;3
profileschemaversion                    1
GroupDocumentsListId                    d174496c-b843-4353-b801-dedb2ea7e25c
disablehelpcollections                   Private
GroupType                                Private
contenttypesusagebackfillversion         3
vti_associatevisitorgroup                4
vti_extenderversion                      16.0.0.25826
GroupAlias                               Spain_Models
LastGroupSitePrivacyUpdated              638777482608801079
vti_associateownergroup                  3
enablehelpcollections                    VCS_EndUser
ProvCorrelationId                       e6278ba1-d0bb-8000-4f3d-967ed84293ce
FollowLinkEnabled                        TRUE
GeneralRetention                         3years
vti_associatemembergroup                 5
GroupDocumentsUrl                        Shared Documents
vti_indexedpropertykeys                  RwbLAG4AZQByAGEAbABSAGUAdABLAG4AdABpAG8ABgA=|
```

Figura 132. Aplicación de propiedad mapeable – DLM

- Búsqueda de Site de SharePoint mediante la query del Adaptive Scope

Tras aplicar la propiedad mapeable, debe de poder buscarse mediante una KQL su `refinable_string:valor` asignado, y conseguir que se liste el Site correspondiente, de esta forma sabemos que los indexadores de Microsoft reconocen la asignación de propiedades nueva:

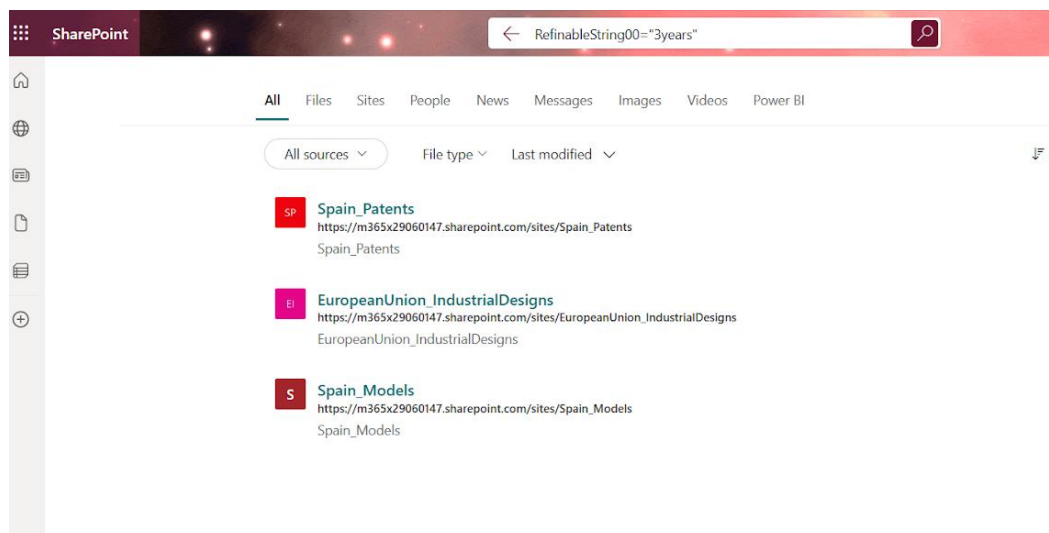


Figura 133. Listado de Sitios de SharePoint consultados – DLM

- Mapeo correcto de Adaptive Scope a los sitios de SharePoint

En los detalles del Adaptive Scope configurado podemos ver en qué sitios de SharePoint se está aplicando, dependiendo de las propiedades que detecta en los meta atributos del sitio.

Vemos que para el Plus_Scope (Scope de plus de un año), se listan todos los sitios deseados:

Adaptive scopes > Plus_Scope

Scope

Plus_Scope

↓ Export 7 items Refresh

Filters: Filters

Display name	Location type
<input type="checkbox"/> https://m365x29060147.sharepoint.com/sites/Spain_Patents	Site
<input type="checkbox"/> https://m365x29060147.sharepoint.com/sites/Spain_Trade...	Site
<input type="checkbox"/> https://m365x29060147.sharepoint.com/sites/EuropeanUnion_Tradem...	Site
<input type="checkbox"/> https://m365x29060147.sharepoint.com/sites/UnitedStates_Trade...	Site
<input type="checkbox"/> https://m365x29060147.sharepoint.com/sites/EuropeanUnion_Industri...	Site
<input type="checkbox"/> https://m365x29060147.sharepoint.com/sites/Spain_Models	Site
<input type="checkbox"/> https://m365x29060147.sharepoint.com/sites/TestGeneralSpecial	Site

Details

Name
Plus_Scope

Description
A scope to extend one year.

Admin unit
Full directory

Type
Site

Query summary
RefinableString01="1year"

Last modified by
Microsoft CDX

Last modified
Mar 20, 2025 8:40 PM

Figura 134. Listado de Sitios de SharePoint, ámbito dinámico– DLM

- Correcta aplicación de las políticas de retención

Existe una herramienta en Microsoft Purview llamada Policy Lookup encargada de listar las políticas de retención que están siendo aplicadas a un Site. Si listamos las políticas del sitio EuropeanUnion_IndustrialDesigns, vemos que están siendo aplicadas dos políticas:

Policy lookup

Search for a specific user, SharePoint site, or Microsoft 365 Group to find out which data lifecycle management policies (retention, label, and auto-labeling) they're included in.

Find policies that include a Enter a site's exact URL

Site Search Clear

Refresh 3 items

Policy name	Scope types	Applications	Last modified	Date created
<input type="checkbox"/> Scope_3years	AdaptiveScope	SharePoint	Mar 22, 2025 6:25 PM	Mar 22, 2025 6:25 PM
<input type="checkbox"/> Policy for plus one year	AdaptiveScope	SharePoint	Mar 23, 2025 8:52 PM	Mar 23, 2025 8:52 PM

Figura 135. Listado de políticas aplicadas a Sitios de SharePoint– DLM

Una es “Policy for plus one year”, que es la política con Scope Adaptativo de un año extra que tienen aplicados todos los sitios de SharePoint, y la otra es “Scope_3years” el cual tiene un nombre auto explicativo, es la política que tiene un Scope Adaptativo de tres años.

Por lo tanto, actualmente el sitio tiene aplicadas las políticas de retención correctas. Guardará por defecto sus documentos durante tres años y si se requiere, se puede asignar un plus de un año más.

11.2.3.1 Juego de pruebas extra

Parte de la solución de Datalifecycle Management la conforman los scripts realizados para el correcto orden de los directorios, subida a SharePoint y configuración de las propiedades, a continuación, se detallan dichas pruebas:

Prueba	Salida	Resultado
extract_fields.py - Documento contiene "Marca nacional" con campo válido	Crea carpeta Spain_Trademarks y subcarpeta con el nombre del cliente	OK
extract_fields.py - Documento sin ninguna palabra clave	Muestra mensaje "No se encontró ningún tipo de interés en el PDF."	OK
extract_fields.py - Campo "Inventor/es:" está mal formateado o ausente	No extrae el campo, no se crea subcarpeta.	OK
create_sites.ps1 - Ejecutado sin errores y Site no existente	Crea nuevo Team Site con bibliotecas Clients y Pending Appointments	OK
create_sites.ps1 - Ejecutado cuando el Site ya existe	Muestra mensaje "El sitio ya existe".	OK
create_sites.ps1 - Biblioteca ya existe	Muestra mensaje "La biblioteca ya existe", sin duplicación.	OK
config_sites.ps1 - CSV contiene URLs y claves válidas	Se conectan los Sites y se escriben las claves sin error en el Property Bag.	OK
config_sites.ps1 - CSV con fila vacía o sin clave	Omite el site, escribe log de "Skipped"	OK
config_sites.ps1 - Módulo PnP no instalado	Mensaje de error y sugerencia de instalación, termina ejecución.	OK
config_sites.ps1 - Clave ya existe y overwrite = false	Registra error en el log de que ya existe, omite modificación.	OK

Tabla 3. Pruebas scripts DLM

11.3 Fase aplicación web

En este apartado se detallarán las pruebas realizadas tanto para el backend como para el frontend de la aplicación web.

11.3.1 Pruebas Backend

Las pruebas realizadas fueron las siguientes:

Prueba	Salida	Resultado
Envío de datos completos y válidos a POST /payment	Genera cookie con JWT, marca cita como pagada, respuesta 200 OK	OK
Envío de UploadDocumentsDTO válido a POST /sharepoint con token JWT	Se generan documentos y se suben correctamente a la biblioteca correspondiente y se devuelve estado 200 OK	OK
Envío correcto de NewEventDTO a POST con token JWT /calendar	Se crea el evento en el calendario de Microsoft 365 y se devuelve estado 200 OK	OK
Envío correcto de UploadDocumentsDTO a POST sin token JWT /sharepoint	JWT inválido detectado por JwtUtils, petición rechazada por TokenFilter, status 403	OK
Llamada a POST /calendar sin token JWT	Filtrado por TokenFilter, se bloquea la petición, respuesta 403	OK
Token manipulado en cookie	JWT inválido detectado por JwtUtils, petición rechazada por TokenFilter, status 403	OK
Se envía un país fuera de los abarcados por la api con un POST al endpoint /sharepoint	Excepción: país no soportado	OK
Se envía un país servicio de los abarcados por la api con un POST al endpoint /sharepoint	Excepción: servicio no soportado	OK

Tabla 4. Pruebas backend

11.3.2 Pruebas Frontend

Las pruebas realizadas para el frontend son las siguientes:

Prueba	Salida	Resultado
Cargar HomePage.tsx	Se muestra correctamente la vista inicial con los servicios legales disponibles	OK
Clic en un ServiceCard	Cambia al siguiente paso (CalendarStep)	OK
CalendarStep.tsx muestra fechas de disponibilidad del empleado	Renderiza correctamente los días con los trabajadores disponibles	OK
Seleccionar franja horaria disponible	Avanza a ClientFormStep y guarda selección	OK
ClientFormStep.tsx muestra formulario completo	Se visualizan los campos requeridos (nombre, email, país, etc.)	OK
Dejar campo obligatorio vacío en ClientFormStep	Desactiva el botón “Continuar” y muestra mensaje de error	OK
Rellenar todos los campos y continuar	Avanza a ConfirmationStep y muestra resumen de la cita y datos introducidos	OK
Confirmar cita desde ConfirmationStep.tsx	Realiza POST al backend, muestra mensaje de éxito	OK
Componente Stepper.tsx refleja progreso correctamente	Marca paso actual con estilo visual	OK

Tabla 5. Pruebas frontend

12 Evaluación de Costos

12.1 Costo del personal

Este trabajo de fin de grado ha combinado dos salidas profesionales: Analista de datos y Desarrollador de Software.

Es un trabajo que está estimado en hacerse en el tiempo que se tardó en desarrollar toda la solución. Es decir, cuatro meses. Los costos por el servicio de dos perfiles de este sector serían los siguientes:

- Analista de datos: 1500€/mes
- Desarrollador de software: 1500€/mes

Total (1500€ * 2 trabajadores * 4) = 12.000€

12.2 Costo del software y hardware

Los usuarios necesitan disponer de licencias Microsoft E5 Compliance, además de tener todo el paquete Office y los servicios para el controlador de dominio y toda la infraestructura de identidad híbrida.

Se usó una calculadora de presupuestos online de Microsoft y dio el total de 905,94€. Este presupuesto es incluyendo 15 usuarios.

El resto de las tecnologías utilizadas son de coste gratuito.

12.3 Costo total

El costo total viene dado por la suma del costo del personal que implementa la solución y el costo de las licencias dando un total de:

Subtotal = 12.000€ + 905,94€ = 12.905,94€.

Total (Incluyendo IVA) = 12.905,94€ + 2.710,25 = 15.616,19€

13 Valoración personal

Este Trabajo de Fin de Grado ha supuesto un desafío técnico y personal que me ha permitido consolidar y aplicar de manera práctica gran parte de los conocimientos adquiridos durante la titulación. La construcción de una solución integral para la protección y gestión de datos me ha llevado a integrar distintas disciplinas: administración de sistemas, seguridad de la información, desarrollo backend y frontend, seguridad en la nube, automatización con scripts, y gobierno de la información.

Desde el punto de vista técnico, he podido enfrentarme a tecnologías reales utilizadas en entornos profesionales, como Microsoft Entra ID, Microsoft Graph, SharePoint, PowerShell, y Microsoft Purview. Esto me ha obligado a salir del enfoque puramente académico dado en la carrera, para entender la lógica, las limitaciones y las posibilidades de estas herramientas en escenarios reales.

Diseñar políticas de retención, clasificaciones automáticas, estrategias de protección mediante etiquetas o detección de riesgos internos ha sido una experiencia formativa profunda. Asimismo, el desarrollo de una aplicación web completa me ha permitido afianzar mis conocimientos en arquitectura de servicios y desarrollo de aplicaciones web, aprendiendo React y Spring Boot.

A nivel personal, este trabajo me ha ayudado a desarrollar autonomía, planificación y capacidad de resolución de problemas. La toma de decisiones de diseño, los errores inesperados, la necesidad de documentar todo correctamente y la integración de tecnologías para proporcionar una solución integral con el mínimo error posible me han hecho más consciente de las exigencias del mundo profesional y de los desafíos que enfrentan las empresas para proteger su información y la importancia que tiene este ámbito en la informática.

En cuanto a los objetivos definidos al inicio del TFG, considero que se han cumplido plenamente. Se ha diseñado e implementado una infraestructura de identidad híbrida, se ha configurado un sistema de protección, retención y clasificación de la información alineado con el contexto de una firma legal, y se ha desarrollado una herramienta web funcional e integrada que demuestra la aplicabilidad de la solución.

Un punto muy importante es señalar que a pesar de que toda la fase de protección de datos ha sido desarrollada gracias a todos los conocimientos que aprendí en un entorno laboral real, hay una gran parte que he tenido que investigar, que sumadas a las demás tecnologías modernas que he aprendido para desarrollar el backend y el frontend, me han permitido desarrollar una solución que es totalmente funcional y aplicable al mundo real.

En conclusión, este TFG no solo ha sido un ejercicio académico, sino una experiencia de desarrollo profesional anticipado, que me ha permitido visualizar con claridad los retos y oportunidades del entorno tecnológico actual y la gran importancia que tiene en las empresas de hoy en día, la protección de la información que maneja.

14 Anexos

El código de la solución implementada se encuentra disponibles en tres repositorios de GitHub diferentes:

- Scripting
[llornn/TFG](#)
- Backend
[llornn/apitfg](#)
- Frontend
[llornn/frontfg](#)

Los repositorios son privados, pero si se recibe una solicitud de revisión se aceptará

15 Bibliografía

Microsoft. (15 de 02 de 2022). *Microsoft Tech Community*. Obtenido de Using Custom SharePoint Site Properties to Apply Microsoft 365 Retention with Adaptive Policy Scopes: <https://techcommunity.microsoft.com/blog/microsoft-security-blog/using-custom-sharepoint-site-properties-to-apply-microsoft-365-retention-with-ad/3133970>

Referencias

- [1] Página Web [Learn about data loss prevention | Microsoft Learn](#) [consulta] 02/03/2025
- [2] Página Web [Guía de instalación de Microsoft Entra Connect y Microsoft Entra Connect Health. - Microsoft Entra ID | Microsoft Learn](#) [consulta] 10/02/2025
- [3] Página Web [Microsoft Purview Information Protection | Microsoft Learn](#) [consulta] 11/03/2025
- [4] Página Web [Adaptive scopes | Microsoft Learn](#) [consulta] 14/03/2025
- [5] Página Web [Register an Entra ID Application to use with PnP PowerShell | PnP PowerShell](#) [consulta] 10/04/2025
- [6] Página Web [Get started with data lifecycle management | Microsoft Learn](#) [consulta] 10/04/2025
- [7] Página Web [Learn about insider risk management | Microsoft Learn](#) [consulta] 12/03/2025
- [8] Página Web [Learn about sensitive information types | Microsoft Learn](#) [consulta] 25/03/2025
- [9] Página Web [Implementing JSON Web Token \(JWT\) Authentication using Spring Security | A Detailed Walkthrough | by Sayan Das | Geek Culture | Medium](#) [consulta] 18/04/2025
- [10] Página Web [Getting Started with Spring Boot: A Comprehensive Tutorial for Beginners | by Marat Miftakhov | The Fresh Writes | Medium](#) [consulta] 19/04/2025
- [11] Página Web [Primeros pasos con React y Typescript | by Lucas Fernández Aragón | Medium](#) [consulta] 08/05/2025
- [12] Página Web [Tutorial - PyMuPDF 1.26.0 documentation](#) [consulta] 14/03/2025
- [13] Página Web [Página principal - EUIPO](#) [consulta] 15/02/2025
- [14] Página Web [Patent Public Search Basic | USPTO](#) [consulta] 15/02/2025
- [15] Página Web [OEPM](#) [consulta] 15/02/2025

Dedicatoria

Por último, me gustaría concluir este trabajo de fin de grado con unos agradecimientos.

Gracias a mis padres, por siempre estar presentes y apoyarme.

Gracias a mis hermanos, por ser una inspiración para mí.

Y Gracias a Rafa, por todos los ánimos dados y por confiar en mí.

Este Trabajo de fin de grado es para todos ustedes, gracias por acompañarme en esta etapa, los quiero mucho.