

# Investigación Operativa mediante fuentes abiertas (OS/INT) a Ebroker Insurance Technologies, S.A.

Trabajo de Fin de Grado

7 de mayo de 2025



ATT&CK<sup>®</sup>



Universidad de Oviedo

**Autor:** Eduardo Blanco Bielsa

**Tutores Corporativos:** Saúl González Llanaea e Iván Buceta Tamargo

**Tutor Académico:** José Manuel Redondo López



## Control de versiones

**Versión actual:** 2025.ES.008

**Fecha:** 07/05/2025

Versión	Fecha	Comentarios de versión
2025.ES.001	25/04/2025	Creación del documento y su plantilla.
2025.ES.002	26/04/2025	Elaboración de capítulos del 1 al 3 incluidos (incluidos anexos).
2025.ES.003	27/04/2025	Elaboración de capítulos del 4 al 6 incluidos.
2025.ES.004	29/04/2025	Elaboración de capítulos del 7 al 10 incluidos.
2025.ES.005	30/04/2025	Primera revisión: corrección de erratas menores y completitud de las explicaciones. Inclusión de comentarios de tutores y contraste corporativos.
2025.ES.006	01/05/2025	Segunda revisión: nueva completitud de las explicaciones mediante comentarios de tutor académico junto con los tutores corporativos.
2025.ES.007	02/05/2025	Tercera revisión: corrección de erratas menores y completitud de la bibliografía consultada, así como de las herramientas empleadas. Revisión del editor de Microsoft Word, con un porcentaje del 100%.
2025.ES.008	07/05/2025	Cuarta revisión: incorporación de sugerencias de tutor académico y revisión del editor de Microsoft Word, con un porcentaje del 100%. Añadida hoja de declaración jurada.

Autor:	Eduardo Blanco Bielsa		© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria			Hoja 1 de 81



# Declaración Responsable

**El alumno:** Eduardo Blanco Bielsa

**Con DNI:** 41012833S

**Y UO:** 285176

## DECLARA

Que esta obra es completamente original y se han citado debidamente las fuentes utilizadas durante la realización de esta.

Y para que conste, lo firma en Oviedo, a 7 de mayo de 2025.

**Firmado:**

Autor:	Eduardo Blanco Bielsa		© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria			Hoja 2 de 81



## Agradecimientos

En primer lugar, me gustaría expresar mi más sincero agradecimiento a mi tutor en Ebroker Saúl González Llanea por su orientación y sus valiosas recomendaciones a lo largo del desarrollo de este Trabajo de Fin de Grado.

Asimismo, deseo agradecer también a mi tutor de la Universidad José Manuel Redondo López por brindarme su dedicación y recursos y por fomentar mi interés por la ciberseguridad y el *Open-Source Intelligence*.

Finalmente, quiero expresar mi más profundo agradecimiento a mi familia, cuyo respaldo y confianza han sido esenciales en mi formación académica y personal. Su apoyo ha sido un pilar fundamental en este camino.

Autor:	Eduardo Blanco Bielsa		© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria			Hoja 3 de 81



## Resumen

En el presente Trabajo de Fin de Grado se aborda una investigación basada en técnicas de *Open-Source Intelligence* (OSINT) para analizar la plantilla de una empresa real y su exposición en fuentes abiertas, incluyendo información disponible en redes sociales y filtraciones de datos, tanto a nivel corporativo como privado. Para ello, se han recopilado y examinado datos obtenidos exclusivamente de fuentes públicas y foros de la Dark web, garantizando en todo momento el cumplimiento de los principios éticos y legales aplicables en materia de ciberseguridad.

Como parte del estudio, se ha desarrollado una herramienta complementaria que permite la visualización gráfica y detallada de la información recopilada mediante el uso de técnicas de *Big Data* para su procesamiento y análisis. Este sistema facilita la identificación de patrones de exposición de datos y posibles vulnerabilidades, proporcionando una visión estructurada y accesible para la toma de decisiones y el cumplimiento de normativas y estándares en materia de seguridad informática.

Los resultados obtenidos evidencian la cantidad de información sensible que puede encontrarse en fuentes abiertas y su potencial impacto en la seguridad de las organizaciones. Este tipo de exposición no solo incrementa el riesgo de sufrir ciberataques, sino que también puede afectar gravemente a la reputación corporativa y, en casos más graves, comprometer la continuidad del negocio. Asimismo, se resalta la importancia de implementar estrategias de concienciación y protección de datos para mitigar los riesgos derivados de la exposición involuntaria de información.

**Palabras clave:** *OSINT* (*Open-Source Intelligence*), fuentes abiertas, ciberataque, *Big Data*, brecha/filtración de datos (*leak*), *Dark web*, *ransomware*, ingeniería social, *MITRE ATT&CK*, huella digital.

Autor:	Eduardo Blanco Bielsa		© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria			Hoja 4 de 81



## Abstract

This Final Degree Project is a research based on Open-Source Intelligence (*OSINT*) techniques to analyze the workforce of a real company and its exposure in open sources, including information available in social networks and data leaks, both at corporate and private levels. To this end, data obtained exclusively from public sources and dark web forums have been collected and examined, ensuring at all times compliance with applicable ethical and legal principles regarding cybersecurity.

As part of the study, a small tool has been developed that allows the graphical and detailed visualization of the information collected through the use of *Big Data* techniques for processing and analysis. This system facilitates the identification of data exposure patterns and possible vulnerabilities, providing a structured and accessible vision for decision making and compliance with regulations and standards in computer security.

The results obtained show the amount of sensitive information that can be found in open sources and its potential impact on the security of organizations. This type of exposure not only increases the risk of cyber-attacks, but can also seriously affect corporate reputation and, in more serious cases, compromise business continuity. It also highlights the importance of implementing awareness and data protection strategies to mitigate the risks arising from unintentional exposure of information.

**Keywords:** *OSINT* (Open-Source Intelligence), open-source, cyber-attack, *Big Data*, data breach/leak, *Dark web*, ransomware, social engineering, *MITRE ATT&CK*, digital footprint.

Autor:	Eduardo Blanco Bielsa			© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008	
Trabajo de Fin de Grado, Convocatoria Ordinaria				Hoja 5 de 81



## Índice de contenido

Control de versiones .....	1
Agradecimientos .....	3
Resumen .....	4
Abstract .....	5
Motivación .....	11
Finalidad del proyecto .....	13
Posibles ámbitos de aplicación .....	15
Identificación de los interesados .....	17
Interesados .....	17
Empresa solicitante: Ebroker Insurance Technologies, S.A. ....	17
Responsables del Proyecto .....	17
Empleados y directivos de Ebroker Insurance Technologies, S.A. ....	17
Comunidad Académica y Profesionales de Ciberseguridad .....	18
Carga de trabajo estimada .....	18
OBS y PBS .....	19
OBS .....	19
PBS .....	20
Informe de Análisis de Exposición .....	21
Herramienta de visualización de datos .....	21
Comparativa de métodos de investigación OSINT .....	21
Guía de buenas prácticas en seguridad y concienciación .....	21
Planificación inicial. WBS .....	22
Riesgos .....	22
Plan de Gestión de riesgos .....	22
Identificación de riesgos .....	22
Registro de riesgos .....	23
Análisis legal complementario y cumplimiento del RGPD .....	25
Presupuesto inicial .....	26
Ejecución del proyecto .....	27

Autor:	Eduardo Blanco Bielsa	© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.	Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria	Hoja 6 de 81	



Plan de Seguimiento de Planificación.....	27
Bitácora de incidencias del proyecto.....	28
Riesgos.....	29
Cierre del proyecto.....	31
Planificación final.....	31
Informe final de riesgos.....	32
Presupuesto final de costes.....	32
Informe de lecciones aprendidas.....	33
Maltego.....	37
BlackBird.....	39
Spiderfoot.....	40
Diferencias clave y valor agregado de este proyecto.....	41
Ejemplos de análisis posibles.....	43
Detección de empleados con mayor exposición.....	44
Relaciones entre empleados y sus perfiles sociales.....	45
Plataformas sociales más frecuentes entre empleados.....	45
Creación del entorno de trabajo.....	47
Configuración de Ubuntu Desktop, Cypher y Jupyter Notebook.....	47
Investigación OSINT.....	48
Solicitud del presupuesto.....	48
Escaneo Activo, Web Scraping y Búsqueda de filtraciones.....	48
Escaneo Activo.....	49
Web Scraping.....	50
Búsqueda de filtraciones.....	51
Consideraciones previas.....	53
Interpretación de los resultados.....	53
Interpretación mediante grafos.....	54
Leyenda del grafo de la investigación.....	54
Interpretación mediante <i>Big Data</i> y visualización de datos.....	63
Empleados agrupados por sus departamentos.....	63

Autor:	Eduardo Blanco Bielsa	© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.	Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria	Hoja 7 de 81	





Plataformas más comunes donde los empleados tienen un perfil social.....	64
Empleados con más perfiles sociales.....	65
Empleados con más brechas de datos .....	66
Discusión .....	67
Hallazgos clave y medidas adoptadas.....	67
Dispositivos comprometidos dentro de la red corporativa.....	67
Implementación obligatoria de autenticación multifactor (MFA) en accesos críticos .....	68
Migración acelerada de controladores de dominio antiguos u obsoletos .....	68
Resumen final de resultados obtenidos .....	69
Conclusiones .....	70
Trabajo futuro .....	71
Difusión de resultados .....	72
Plan de Gestión de riesgos.....	74
Glosario de términos.....	76
Herramientas utilizadas .....	79

Autor:	Eduardo Blanco Bielsa	© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.	Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria	Hoja 8 de 81	



## Índice de tablas

Tabla 1. Carga de trabajo estimada.....	19
Tabla 2. OBS. ....	20
Tabla 3. Riesgos identificados. ....	30
Tabla 4. Resumen final de resultados obtenidos.....	69
Tabla 5. Plan de Gestión de riesgos.....	75

Autor:	Eduardo Blanco Bielsa	© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.	Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria	Hoja 9 de 81	



## Índice de ilustraciones

Ilustración 1. Evolución de hechos conocidos por categorías delictivas, INCIBE.....	11
Ilustración 2. Principal método de infección de un ransomware, Statista. ....	13
Ilustración 3. Marco OSINT Framework, primer nivel.....	16
Ilustración 4. Planificación inicial, WBS. ....	22
Ilustración 5. Planificación final. ....	31
Ilustración 6. Maltego. ....	38
Ilustración 7. BlackBird, CLI. ....	39
Ilustración 8. Spiderfoot. ....	41
Ilustración 9. Detección de empleados con mayor exposición, formato treemap. ....	44
Ilustración 10. Detección de empleados con mayor exposición, formato scatter. ....	44
Ilustración 11. Relaciones entre empleados y sus perfiles sociales, formato treemap. ....	45
Ilustración 12. Plataformas sociales más frecuentes entre empleados, formato bar. ..	46
Ilustración 13. Plataformas sociales más frecuentes entre empleados, formato treemap. ....	46
Ilustración 14. Diagrama de flujo de la metodología empleada en la investigación. ....	52
Ilustración 15. Grafo completo de la investigación.....	55
Ilustración 16. Parte ampliada del grafo completo de la investigación. ....	56
Ilustración 17. Grafo de los empleados con sus respectivos departamentos de trabajo. ....	57
Ilustración 18. Grafo de empleados junto con todos sus perfiles sociales. ....	58
Ilustración 19. Parte ampliada del grafo de empleados junto con todos sus perfiles sociales.....	59
Ilustración 20. Grafo de empleados con todas sus filtraciones de datos.....	60
Ilustración 21. Parte ampliada del grafo de empleados con todas sus filtraciones de datos. ....	61
Ilustración 22. Grafo con todos los datos de la investigación sobre un empleado concreto.....	62
Ilustración 23. Gráfico de empleados agrupados por sus departamentos, formato bar. ....	63
Ilustración 24. Gráfico de plataformas sociales más usadas por los empleados, formato bar. ....	64
Ilustración 25. Gráfico de plataformas sociales más usadas por los empleados, formato treemap. ....	64
Ilustración 26. Gráfico de empleados con más perfiles sociales, formato treemap. ....	65
Ilustración 27. Gráfico de empleados con más brechas de datos, formato scatter. ....	66
Ilustración 28. Gráfico de empleados con más brechas de datos, formato treemap....	66

Autor:	Eduardo Blanco Bielsa	© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.	Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria	Hoja 10 de 81	

# Capítulo 1. Introducción

## Motivación

Este Trabajo de Fin de Grado surge de la necesidad identificada por la empresa Ebroker Insurance Technologies, S.A. de evaluar su nivel de exposición en fuentes abiertas para determinar los riesgos asociados a la disponibilidad de información sensible en el entorno digital. En un contexto donde los ciberataques, como campañas de *ransomware* o fraudes mediante ingeniería social, se han convertido en amenazas recurrentes para las organizaciones, resulta imprescindible comprender cómo la información accesible públicamente puede ser utilizada por actores malintencionados para comprometer la seguridad corporativa.

A continuación, se incluye una infografía del INCIBE donde se indican las amenazas delictivas más comunes hoy en día, siendo la más recurrente el fraude informático:

HECHOS CONOCIDOS	2019	2020	2021	2022	2023
ACCESO E INTERCEPTACIÓN ILÍCITA	4.004	4.653	5.342	5.578	7.367
AMENAZAS Y COACCIONES	12.782	14.066	17.319	15.982	17.472
CONTRA EL HONOR	1.422	1.550	1.426	1.191	1.174
CONTRA PROPIEDAD INDUST./INTELEC.	197	125	137	114	64
DELITOS SEXUALES(*)	1.774	1.783	1.628	1.646	1.804
FALSIFICACIÓN INFORMÁTICA	4.275	6.289	10.476	12.569	15.137
FRAUDE INFORMÁTICO	192.375	257.907	267.011	335.995	427.448
INTERFERENCIA DATOS Y EN SISTEMA	1.473	1.590	2.138	1.662	1.659
<b>Total HECHOS CONOCIDOS</b>	<b>218.302</b>	<b>287.963</b>	<b>305.477</b>	<b>374.737</b>	<b>472.125</b>

(\*)Excluidos las agresiones sexuales con/sin penetración y los abusos sexuales con penetración

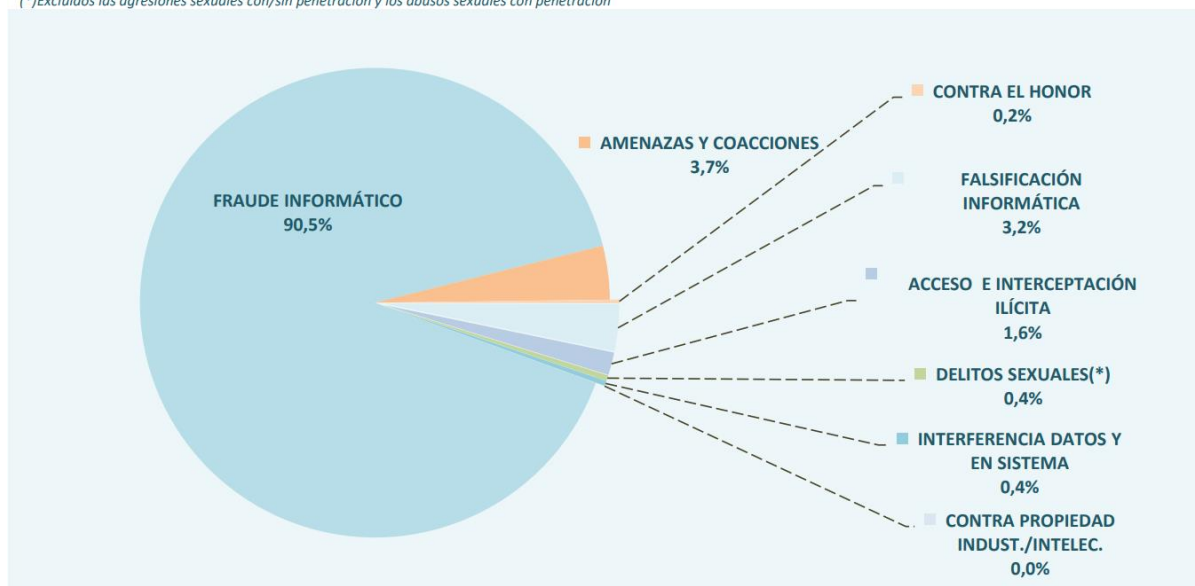


Ilustración 1. Evolución de hechos conocidos por categorías delictivas, INCIBE.

Autor:	Eduardo Blanco Bielsa	© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.	Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria	Hoja 11 de 81	



Diversos informes oficiales del Ministerio del Interior y de cuerpos policiales reflejan el preocupante aumento de incidentes relacionados con ransomware y filtraciones de datos. Estas fuentes destacan cómo los ciberdelincuentes utilizan información obtenida de fuentes abiertas para lanzar ataques dirigidos contra empresas, afectando tanto a su operatividad como a su reputación. La tendencia ascendente en estos ataques pone de manifiesto la necesidad de adoptar estrategias proactivas para mitigar los riesgos asociados a la exposición digital.

Ebroker es una empresa de tipo insurtech, que opera en el sector seguros, el cual es un sector empresarial muy regulado y que está sujeto tanto a normativas de ámbito nacional como europeo. El pasado 17 de enero de 2025 comenzó a ser obligatorio la aplicación del [reglamento DORA](#), una normativa publicada en 2023 que tiene como principal objetivo mejorar la resiliencia operativa y la ciberseguridad en las organizaciones financieras (las compañías de seguros se consideran organizaciones financieras), con lo cual este estudio ayudará también al cumplimiento de dicha norma, mejorando el nivel de ciberseguridad de la empresa y auditando el nivel de exposición de sus empleados.

Con este estudio, se pretende analizar en detalle la cantidad y el tipo de datos que pueden obtenerse mediante técnicas de *OSINT*, identificando posibles vulnerabilidades derivadas de la exposición de la información de la plantilla de Ebroker Insurance Technologies, S.A. en redes sociales, foros y otras fuentes públicas, incluyendo la *Dark web*. La investigación no sólo busca evidenciar estos riesgos, sino también aportar soluciones para mitigar su impacto.

Para ello, se ha desarrollado una herramienta basada en Big Data que permite visualizar de forma gráfica los datos obtenidos en la investigación de una forma similar a soluciones tradicionales. Este sistema permite evaluar de manera estructurada la información expuesta y facilita la toma de decisiones en materia de ciberseguridad.

En un panorama donde las amenazas digitales evolucionan constantemente, estudios como el presente resultan fundamentales para fortalecer las estrategias de protección de las empresas. Con esta investigación se espera contribuir a la concienciación sobre la importancia de minimizar la exposición digital y reforzar las medidas de seguridad necesarias para prevenir ataques informáticos.

Autor:	Eduardo Blanco Bielsa		© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria			Hoja 12 de 81

## Finalidad del proyecto

Este proyecto tiene como principal objetivo proporcionar a la empresa Ebroker Insurance Technologies, S.A. un análisis detallado de su nivel de exposición en fuentes abiertas, permitiéndole conocer los riesgos asociados a la disponibilidad de información sensible y tomar decisiones fundamentadas para mejorar su seguridad. Para ello se ha aplicado una metodología basada en el marco *MITRE ATT&CK*, ampliamente reconocido en el ámbito de la ciberseguridad por su enfoque estructurado en la identificación de tácticas, técnicas y procedimientos utilizados por actores malintencionados.

Además del análisis de vulnerabilidades, el proyecto busca concienciar a la plantilla sobre los riesgos asociados a la exposición de información en entornos digitales. Muchos ataques actuales, como campañas de ransomware o fraudes de ingeniería social, se basan en datos obtenidos de fuentes abiertas, por lo que sensibilizar a los empleados sobre la protección de su información es una medida clave para la prevención de los incidentes de seguridad.

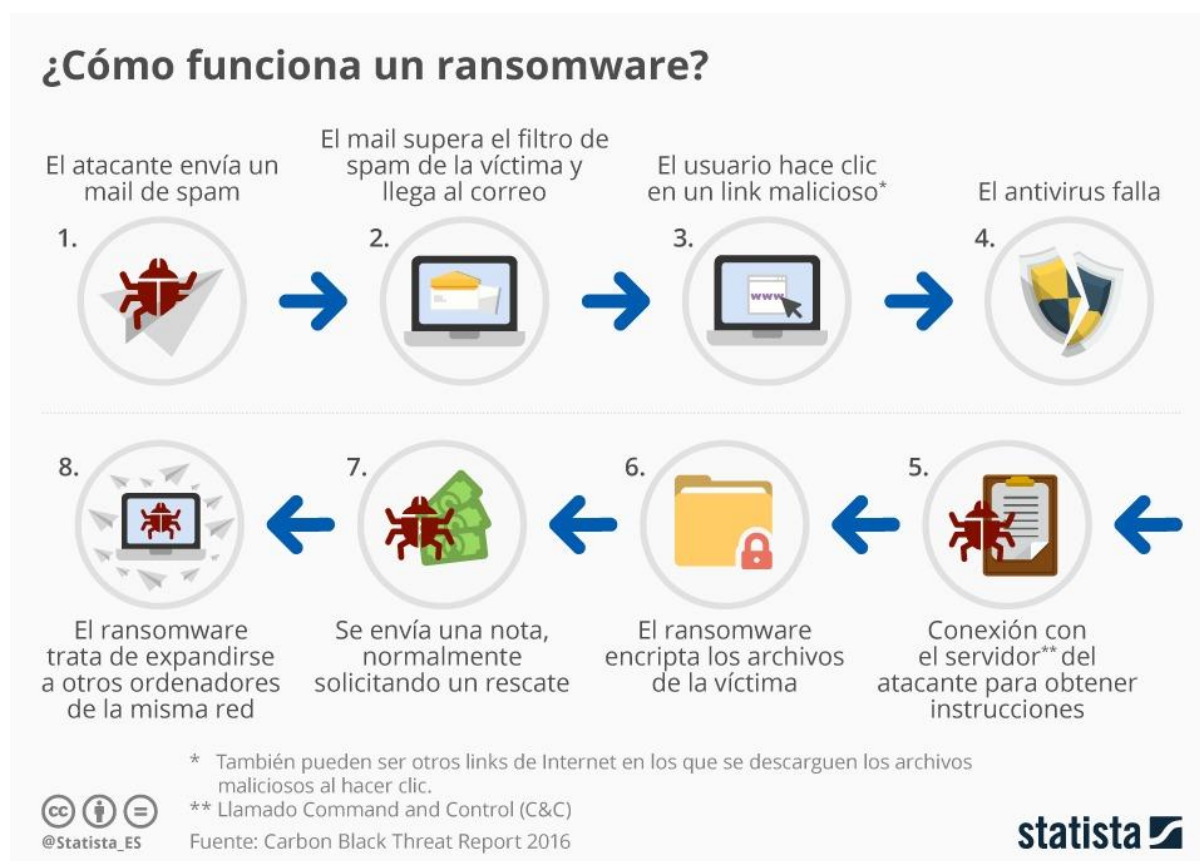


Ilustración 2. Principal método de infección de un ransomware, Statista.

Autor:	Eduardo Blanco Bielsa	© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.	Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria	Hoja 13 de 81	



Otro de los pilares fundamentales de esta investigación es la evaluación comparativa con herramientas OSINT tradicionales, como Maltego. Se busca demostrar que, aunque Maltego es una solución ampliamente utilizada, sus datos no siempre son veraces debido a su automatización en la recopilación de información, lo que suele generar “ruido” en los resultados finales. En contraste, la metodología empleada en este proyecto garantiza la precisión de los datos obtenidos, incluso si el proceso requiere más tiempo. Además, la investigación resalta la reducción significativa de los costes asociados a este tipo de análisis, ofreciendo a Ebroker Insurance Technologies, S.A. una solución más eficiente y adaptada a sus necesidades.

En definitiva, este proyecto no sólo permitirá a Ebroker Insurance Technologies, S.A. evaluar y mejorar su postura en ciberseguridad, sino que también contribuirá al desarrollo de una metodología replicable en otras organizaciones. Al basarse en el marco *MITRE ATT&CK*, se asegura un enfoque estructurado y alineado con estándares reconocidos en la detección y mitigación de amenazas digitales.

Autor:	Eduardo Blanco Bielsa		© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria			Hoja 14 de 81



# Capítulo 2. Fijación de Objetivos

## Posibles ámbitos de aplicación

El procedimiento empleado en esta investigación es extrapolable a un amplio abanico de entidades, incluyendo empresas, corporaciones, gobiernos, instituciones y colectivos, así como a nivel individual. La metodología basada en *OSINT* y el marco *MITRE ATT&CK* permite adaptar el análisis de exposición digital a distintos contextos, proporcionando una evaluación precisa de los riesgos asociados a la presencia de información en fuentes abiertas.

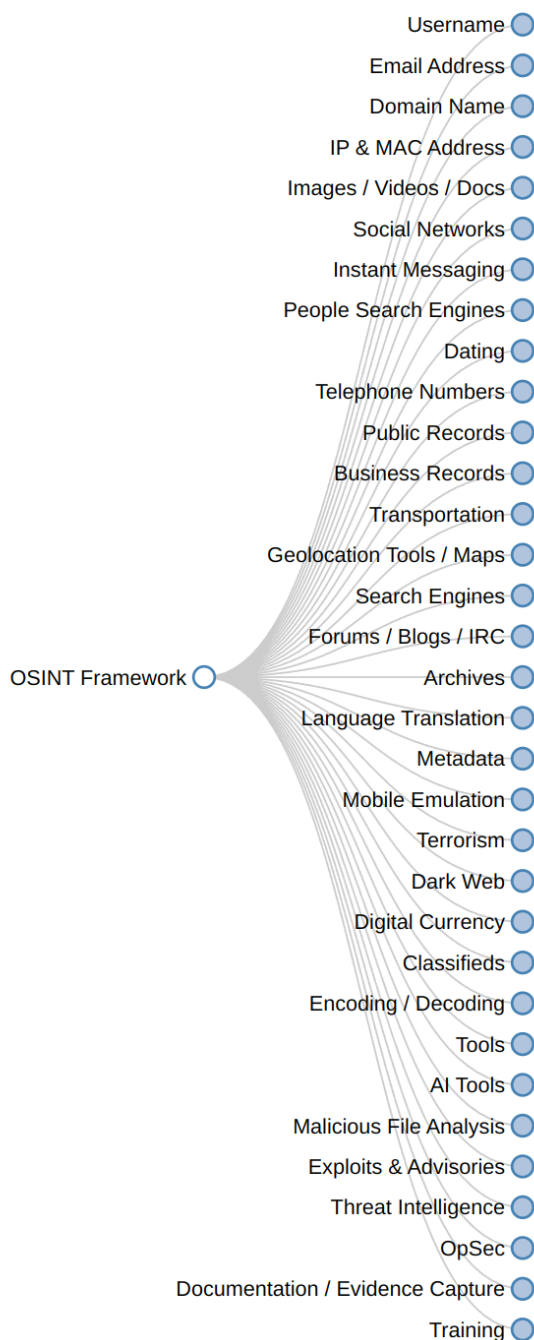
En el ámbito empresarial, este procedimiento puede aplicarse a organizaciones de cualquier sector que deseen conocer su nivel de exposición digital y fortalecer sus estrategias de ciberseguridad. La creciente digitalización y el uso masivo de plataformas en línea han incrementado la cantidad de datos accesibles públicamente, lo que hace imprescindible la implementación de medidas preventivas para reducir posibles vulnerabilidades.

A nivel gubernamental e institucional, el análisis *OSINT* puede emplearse para evaluar la seguridad de entidades públicas y sus integrantes, ayudando a identificar posibles brechas de información que puedan ser explotadas por actores malintencionados. En este contexto, la protección de datos sensible es crucial para garantizar la seguridad nacional y la integridad de infraestructuras críticas.

Asimismo, este procedimiento puede aplicarse en el ámbito personal, permitiendo a individuos evaluar su propia huella digital y minimizar su exposición a posibles amenazas, como el robo de identidad o la ingeniería social. La concienciación sobre la privacidad en línea es un aspecto clave en un entorno donde la información personal se encuentra cada vez más expuesta.

Autor:	Eduardo Blanco Bielsa		© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria			Hoja 15 de 81





*Ilustración 3. Marco OSINT Framework, primer nivel.*

En definitiva, la metodología desarrollada en esta investigación puede ser implementada en cualquier escenario donde la seguridad de la información sea una prioridad o existan regulaciones específicas en materia de ciberseguridad que requieran este tipo de actuaciones. Su capacidad de adaptación a distintos sectores y niveles permite que los resultados obtenidos sean relevantes para la toma de decisiones estratégicas en materia de ciberseguridad.

Autor:	Eduardo Blanco Bielsa		© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria			Hoja 16 de 81



# Capítulo 3. Planificación y Gestión del TFG

## Identificación de los interesados

A lo largo del desarrollo de esta investigación, se han identificado diversos interesados que desempeñan un papel clave en la aplicación y el impacto de los resultados obtenidos.

### Interesados

#### Empresa solicitante: Ebroker Insurance Technologies, S.A.

- **Tipo:** empresa privada del sector tecnológico asegurador.
- **Intereses:** evaluar el nivel de exposición de su plantilla, mejorar su estrategia de ciberseguridad y optimizar sus protocolos internos de protección de datos.
- **Involucración:** revisión de los resultados, implementación de recomendaciones y aplicación de medidas correctivas en función del informe de análisis.

#### Responsables del Proyecto

- **Investigador (Eduardo Blanco Bielsa):** encargado del desarrollo de la metodología, la obtención de datos mediante técnicas *OSINT*, la implementación de la herramienta de visualización y la redacción del informe final.
- **Tutor académico (José Manuel Redondo López):** brinda orientación metodológica, revisa los avances y supervisa la alineación del trabajo de acuerdo con los estándares académicos.
- **Tutor profesional (Saúl González Llaneza):** proporciona asesoramiento en las prácticas de análisis realizadas, así como la visualización de los datos obtenidos y supervisa la alineación del trabajo de acuerdo con los estándares Ebroker Insurance Technologies, S.A.

#### Empleados y directivos de Ebroker Insurance Technologies, S.A.

- **Intereses:** comprender los riesgos asociados a la exposición de información en fuentes abiertas y aplicar las mejores prácticas en ciberseguridad.

Autor:	Eduardo Blanco Bielsa			© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo		Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria				Hoja 17 de 81



- **Involucración:** participación en una sesión de concienciación y aplicación de medidas de seguridad sugeridas en la guía de buenas prácticas.

## Comunidad Académica y Profesionales de Ciberseguridad

- **Intereses:** evaluar la metodología propuesta y su aplicabilidad en otras investigaciones *OSINT*.
- **Involucración:** posible interés en replicar o mejorar el enfoque realizado en este trabajo.

## Carga de trabajo estimada

La ejecución del proyecto se ha planificado para cubrir un total de **338 horas**, distribuidas de la siguiente manera:

Tarea	Horas estimadas
Inicio de la investigación (reuniones y definición del alcance).	5h
Configuración del entorno.	12h
Investigación <i>OSINT</i> .	170h
Validación de datos.	38h
Desarrollo de herramientas.	55h
Pruebas y optimización de la herramienta (depuración, mejoras y ajustes finales).	18h
Elaboración de informe detallado corporativo.	8h
Elaboración de la guía de buenas prácticas.	3h
Exposición de los hallazgos de forma corporativa.	1h
Redacción y revisión del TFG.	28h



Tabla 1. Carga de trabajo estimada.

Este desglose asegura que se cumple con los requisitos establecidos y cada tarea recibe la dedicación necesaria para alcanzar los objetivos definidos.

## OBS y PBS

### OBS

A continuación, se define la asignación de tareas a los distintos recursos de la empresa:

Nombre de la Tarea	Nombres de los recursos
<b>Inicio de la investigación</b>	
Reunión de arranque de la investigación.	Analista de ciberseguridad, CTO.
<b>Configuración del entorno.</b>	
Configuración de Ubuntu.	Analista de ciberseguridad.
Configuración BBDD Cypher.	Analista de ciberseguridad.
Configuración del entorno virtual de Python y librerías asociadas.	Analista de ciberseguridad.
<b>Investigación OSINT.</b>	
Solicitud de presupuesto.	Analista de ciberseguridad, director de negocio.
Escaneo Activo.	Analista de ciberseguridad.
Web Scraping.	Analista de ciberseguridad.
Búsqueda de filtraciones.	Analista de ciberseguridad.



Validación de datos.	Analista de ciberseguridad.
<b>Desarrollo de herramientas.</b>	
Creación de Jupyter Notebook.	Analista de ciberseguridad.
Creación de grafo visual.	Analista de ciberseguridad.
Creación de gráficos visuales.	Analista de ciberseguridad.
Pruebas y optimización de la herramienta.	Analista de ciberseguridad.
<b>Transición.</b>	
Elaboración del informe corporativo (redacción formal de hallazgos).	Analista de ciberseguridad.
Elaboración de una guía de buenas prácticas para empleados.	Analista de ciberseguridad.
Exposición corporativa de hallazgos y concienciación.	Analista de ciberseguridad.
<b>Cierre de la investigación.</b>	

Tabla 2. OBS.

## PBS

La presente investigación tiene como resultado una serie de productos clave que permitirán a la empresa Ebroker Insurance Technologies, S.A. y a otras organizaciones comprender su nivel de exposición en fuentes abiertas y mejorar sus estrategias de ciberseguridad. Se incluyen a continuación dichos productos.

Autor:	Eduardo Blanco Bielsa	© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.	Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria	Hoja 20 de 81	



## Informe de Análisis de Exposición

- Documento detallado que recopila la información obtenida mediante técnicas *OSINT*, identificando la exposición de datos de la plantilla de Ebroker Insurance Technologies, S.A. en fuentes públicas, redes sociales y foros de la Dark web.
- Evaluación del nivel de riesgo asociado a la información recopilada.
- Recomendaciones para mitigar los riesgos detectados y mejorar la seguridad de la empresa.

## Herramienta de visualización de datos

- Aplicación complementaria basada en *Big Data* que permite procesar y visualizar los datos obtenidos de forma estructurada y detallada.
- Implementación de funcionalidades para analizar patrones y correlaciones entre los datos recopilados.

## Comparativa de métodos de investigación OSINT

- Análisis de las diferencias entre el enfoque utilizado en esta investigación y herramientas comerciales como Maltego.
- Evaluación del balance entre precisión, veracidad de los datos y costes operativos.
- Justificación de la elección de la metodología basada en el marco MITRE ATT&CK.

## Guía de buenas prácticas en seguridad y concienciación

- Documento destinado a la plantilla de Ebroker Insurance Technologies, S.A. con recomendaciones para minimizar la exposición de datos en fuentes abiertas.
- Estrategias para mejorar la seguridad en el ámbito corporativo.
- Medidas preventivas contra amenazas como *ransomware*, *phishing* e ingeniería social.

Estos productos finales no sólo permitirán a Ebroker Insurance Technologies, S.A. mejorar su postura en ciberseguridad, sino que también servirán como referencia para la aplicación de metodologías *OSINT* en otros sectores y organizaciones.

Autor:	Eduardo Blanco Bielsa		© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria			Hoja 21 de 81

## Planificación inicial. WBS

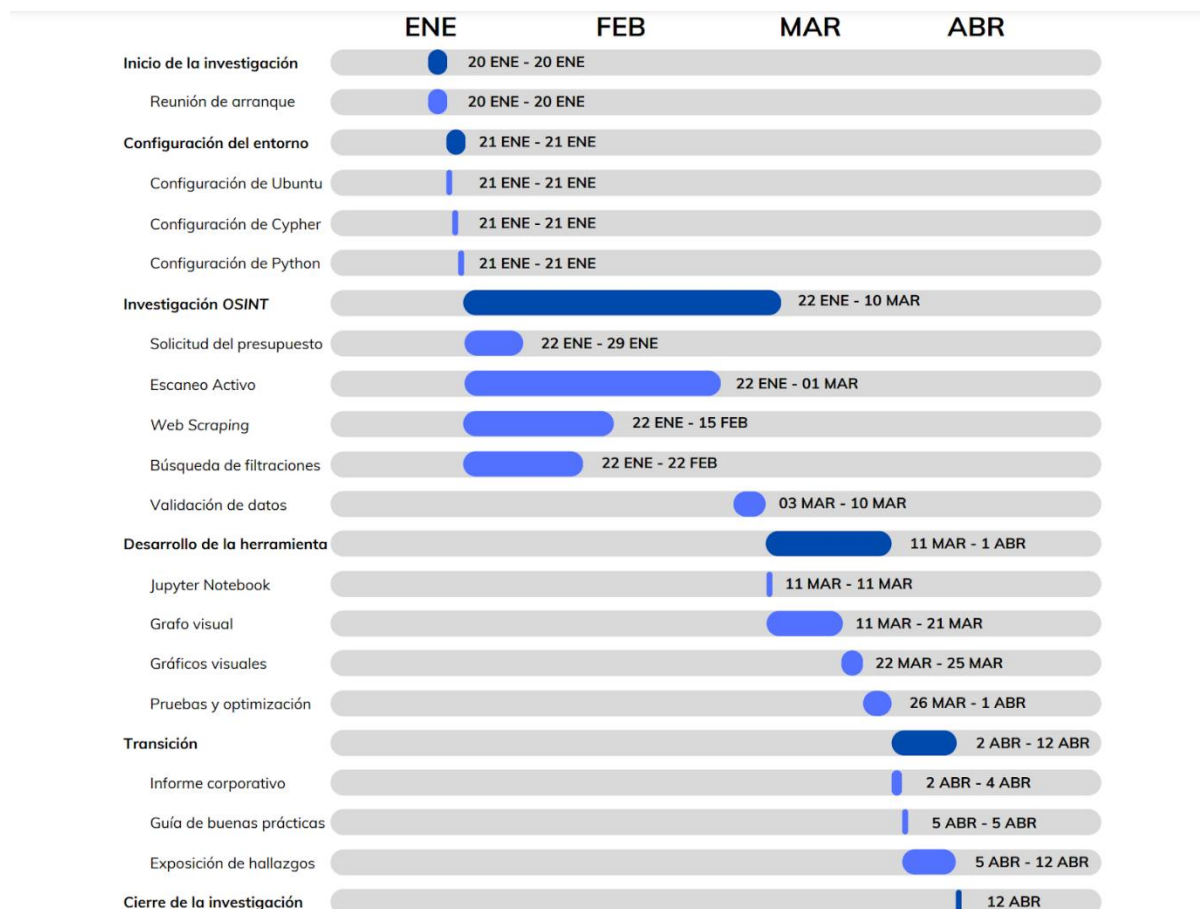


Ilustración 4. Planificación inicial, WBS.

## Riesgos

### Plan de Gestión de riesgos

[Enlace al Plan de Gestión de Riesgos.](#)

### Identificación de riesgos

Durante el desarrollo de la investigación, se han identificado diversos riesgos que pueden comprometer la obtención, el uso y la integridad de los datos. Estos riesgos se dividen en varias categorías, incluyendo aspectos legales, éticos y técnicos. A continuación, se identifican los riesgos asociados a la metodología empleada:

- Acceso a información privada sin consentimiento explícito.
- Cuestiones legales y de consentimiento.

Autor:	Eduardo Blanco Bielsa	© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.	Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria	Hoja 22 de 81	



- Dificultades en la anonimización de datos.
- Reacciones adversas de los sujetos investigados.
- Uso indebido de la información recopilada.
- Limitaciones técnicas en la recopilación de datos (*Web Scraping*).
- Exposición a *malware* en foros de la *Dark web*.
- Riesgo de phishing en entornos de ciberdelincuencia.
- Riesgo de trazabilidad y desanonimización del investigador.
- Fiabilidad y veracidad de las fuentes consultadas.
- Cambio constante en las políticas de las plataformas analizadas.
- Bloqueo de cuentas utilizadas en la investigación.
- Posibilidad de no encontrar filtraciones relevantes.
- Ausencia de perfiles sociales vinculados a los sujetos investigados.
- Baja exposición de información personal en perfiles analizados.

## Registro de riesgos

Durante el desarrollo de la investigación, se han identificado diversos riesgos que pueden comprometer la obtención, el uso y la integridad de los datos. Estos riesgos se dividen en varias categorías, incluyendo aspectos legales, éticos y técnicos. A continuación, se describen los riesgos asociados a la metodología empleada:

- **Acceso a información privada sin consentimiento explícito:** aunque la investigación se basa en datos públicos, existe el riesgo de acceder a información que, si bien está disponible en la red, podría considerarse privada según normativas de privacidad y protección de datos.
- **Cuestiones legales y de consentimiento:** la recopilación y el análisis de información de terceros pueden entrar en conflicto con regulaciones de protección de datos, como el RGPD, especialmente si no se encuentra con un consentimiento explícito de las personas investigadas.
- **Dificultades en la anonimización de datos:** asegurar la anonimización efectiva de los datos obtenidos puede ser complejo, lo que podría exponer a individuos o entidades si la información recopilada no se procesa adecuadamente.

Autor:	Eduardo Blanco Bielsa		© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria			Hoja 23 de 81





- **Reacciones adversas de los sujetos investigados:** la identificación de información personal o corporativa podría generar preocupaciones entre las personas afectadas, provocando posibles conflictos legales o de preocupación.
- **Uso indebido de la información recopilada:** existe el riesgo de que los datos obtenidos sean robados y utilizados con fines malintencionados, lo que podría derivar en la criminalización de la investigación o en su empleo para actividades ilícitas.
- **Limitaciones técnicas en la recopilación de datos (*Web Scraping*):** la implementación de técnicas de *scraping* puede verse obstaculizada por medidas de seguridad, como *Captchas*, autenticación en dos pasos (2FA) o sistemas avanzados de detección de *bots*.
- **Exposición a *malware* en foros de la *Dark web*:** durante la consulta de fuentes en la *Dark web*, existe el riesgo de exposición a malware, lo que podría comprometer la seguridad del dispositivo empleado en la investigación.
- **Riesgo de phishing en entornos de ciberdelincuencia:** la interacción con plataformas de compraventa de filtraciones o foros clandestinos podría exponer al investigador a ataques de *phishing* dirigidos, comprometiendo credenciales e información sensible.
- **Riesgo de trazabilidad y desanonimización del investigador:** el acceso a foros en la *Dark web* conlleva el peligro de que la identidad del investigador sea rastreada si no se aplican medidas adecuadas de anonimización.
- **Fiabilidad y veracidad de las fuentes consultadas:** no toda la información disponible en fuentes OSINT es precisa o actualizada, lo que puede generar falsos positivos o conclusiones erróneas.
- **Cambio constante en las políticas de las plataformas analizadas:** redes sociales y foros modifican constantemente sus términos de uso, además de que algunas plataformas desaparecen y surgen otras nuevas. Estos cambios pueden retrasar o limitar la efectividad de la investigación, afectando la disponibilidad y accesibilidad de la información.
- **Bloqueo de cuentas utilizadas en la investigación:** la ejecución de consultas automatizadas o el acceso reiterado a perfiles de usuarios podría desencadenar restricciones o bloqueos de las cuentas empleadas en la investigación.
- **Posibilidad de no encontrar filtraciones relevantes:** a pesar de los esfuerzos en la búsqueda de información, es posible que la investigación no detecte

Autor:	Eduardo Blanco Bielsa		© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria			Hoja 24 de 81



ninguna filtración de datos significativa, lo que podría afectar a los resultados esperados.

- **Ausencia de perfiles sociales vinculados a los sujetos investigados:** si los individuos analizados no disponen de una huella digital visible en redes sociales, la recolección de información relevante se vería limitada.
- **Baja exposición de información personal en perfiles analizados:** aunque los sujetos analizados tengan presencia en redes sociales, podrían haber configurado su privacidad de manera estricta, limitando el acceso a información de interés.

## Análisis legal complementario y cumplimiento del RGPD

Desde la fase de diseño, el presente trabajo ha considerado los posibles riesgos legales asociados a la obtención y tratamiento de información pública relativa a personas físicas. Aunque toda la información procesada ha sido obtenida a través de fuentes abiertas, accesibles libremente en internet, se ha procurado alinear el desarrollo del proyecto con los principios y obligaciones establecidos en el **Reglamento (UE) 2016/679, General de Protección de Datos (RGPD)**.

En particular, se ha aplicado el **principio de minimización de datos** recogido en el **artículo 5.1.c del RGPD**, limitando el tratamiento exclusivamente a aquellos datos estrictamente necesarios para alcanzar el objetivo de la investigación: evaluar el nivel de exposición pública de los perfiles vinculados a la organización, en el contexto de la ciberseguridad corporativa.

Los datos personales tratados han sido aquellos que los propios interesados han hecho públicos de forma manifiesta, como por ejemplo en redes sociales, foros o sitios web corporativos. Este enfoque se encuentra amparado por el **considerando 26 del RGPD**, que establece que el tratamiento de datos no se considerará aplicable cuando estos no puedan vincularse de forma razonable a una persona física identificada o identificable.

La **base jurídica** que sustenta dicho tratamiento es el **interés legítimo del responsable del tratamiento** (artículo 6.1.f del RGPD), dado que el propósito del análisis es preventivo y está orientado a mejorar la seguridad interna de la organización, así como a mitigar posibles riesgos derivados de la exposición involuntaria de información.

Asimismo, se han adoptado medidas técnicas orientadas a la **protección de la privacidad**, como la **anonimización** y la **seudonimización** de los datos analizados, garantizando que la información no pueda ser utilizada para identificar de forma directa

Autor:	Eduardo Blanco Bielsa	© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.	Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria	Hoja 25 de 81	



a ninguna persona física. En ningún momento se han llevado a cabo procesos de **perfilado automatizado** ni se han tomado decisiones individuales basadas en los datos recopilados.

Por último, se reconoce que, en el caso de una aplicación recurrente o a gran escala de este tipo de análisis dentro de una organización, se podrían generar implicaciones jurídicas adicionales en materia de protección de datos. En este sentido, se recomienda que cualquier implantación futura en un entorno de producción vaya acompañada de una **Evaluación de Impacto relativa a la Protección de Datos (DPIA)**, conforme a lo previsto en el **artículo 35 del RGPD**.

## Presupuesto inicial

Para la realización de esta investigación, se estimó un presupuesto inicial de **50,00 euros**, destinado principalmente a la contratación de servicios específicos para el análisis *OSINT* y la adquisición de herramientas complementarias.

La totalidad del presupuesto se gestionó mediante criptomonedas, concretamente *Bitcoin*, debido a la naturaleza de las plataformas y herramientas utilizadas, las cuales operan en entornos donde esta forma de pago es predominante por razones de privacidad y seguridad. Según la tasa de cambio del **24 de febrero de 2025**, donde **1 BTC** equivalía a **90.508,3 euros**, este presupuesto correspondía exactamente a **0,000552 BTC**. Es importante destacar que, al realizar pagos en estas plataformas, se aplicó una comisión por transacción, habitual en este tipo de operaciones, que supuso un coste adicional aproximado de **0,0001 BTC (0,90 euros)**. Este importe se deriva del uso del monedero de las comisiones cobradas por el proveedor de la cartera digital (Binance) al realizar los pagos con Bitcoin a proveedores de herramientas *OSINT*.

Adicionalmente, aunque el proyecto se ha desarrollado utilizando recursos personales y gratuitos en su mayoría, se puede presentar un **presupuesto ampliado orientativo** para reflejar el valor real de los recursos empleados:

- **Medios humanos:**
  - El investigador dedicará aproximadamente **338 horas** al desarrollo completo del proyecto (investigación, recopilación de datos, análisis, desarrollo de herramientas, redacción del informe y revisión), lo que, considerando una tarifa media de **35,00 euros/hora**, supondría un coste estimado de **11830,00 euros**.

Autor:	Eduardo Blanco Bielsa		© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria			Hoja 26 de 81

- El director de tecnología (CTO) dedicará aproximadamente **10 horas** a la lectura, corroboración y construcción de la investigación, lo que supondría un coste estimado de **800,00 euros** adicionales, considerando una tarifa media de **80,00 euros/hora**.
- Estas tarifas se corresponden con las tarifas generalmente empleadas en la empresa solicitante.
- **Medios materiales:**
  - **Equipo informático:** ordenador portátil personal de gama media-alta (**1479,00 euros**).
  - **Conexión a Internet:** asumida como recurso doméstico (**3 × 38 euros = 114,00 euros**).
  - **VPN:** se utilizó una versión profesional de pago de NordVPN de **98,00 euros/año**.

Dado que la empresa Ebroker Insurance Technologies, S.A. es tanto el sujeto de estudio como la solicitante del análisis, no se ha considerado necesario crear un apartado de “Presupuesto de costes” y otro de “Presupuesto de Cliente”.

## Ejecución del proyecto

### Plan de Seguimiento de Planificación

Para garantizar el cumplimiento de los objetivos del proyecto dentro del tiempo estimado, se implementó un sistema de seguimiento basado en la documentación y la gestión eficiente de tareas.

Se utilizó **Obsidian** como herramienta principal para la redacción y estructuración de la información recopilada. A lo largo de la investigación se llevó a cabo un control iterativo de los avances, asegurando que cada fase del proyecto se completase antes de dar paso a la siguiente.

Las tareas se organizaron en un flujo de trabajo que permitía evaluar constantemente el progreso y realizar ajustes en la planificación si era necesario. Los principales puntos de seguimiento incluyeron:

- Registro detallado de cada búsqueda y hallazgo relevante mediante *OS/INT*.
- Análisis de tiempos dedicados a cada iteración del proceso de investigación.

Autor:	Eduardo Blanco Bielsa		© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria			Hoja 27 de 81



- Evaluación del volumen y relevancia de los datos obtenidos en cada ciclo.
- Registro de dificultades técnicas y estrategias de resolución.
- Seguimiento del desarrollo de la herramienta en Python, con pruebas y mejoras progresivas.

Este enfoque permitió mantener un control efectivo sobre la evolución del proyecto, asegurando la optimización de los recursos y la finalización de cada fase dentro del marco de tiempo establecido.

## Bitácora de incidencias del proyecto

A lo largo de la ejecución del proyecto, no se han registrado incidencias críticas que comprometiesen el desarrollo de la investigación. No obstante, se han identificado algunos obstáculos que requirieron ajustes metodológicos para garantizar la continuidad del análisis.

En primer lugar, algunas plataformas implementaron restricciones temporales y bloqueos en las cuentas utilizadas para la recolección de la información, lo que obligó a emplear distintos perfiles y estrategias de acceso. Asimismo, durante la fase de recopilación de datos, ciertos sitios web incorporaron protecciones avanzadas contra el *web scraping*, como *captchas* recurrentes y sistemas de detección de actividad automatizada, lo que dificultó la extracción de información.

Otro desafío estuvo relacionado con el acceso a foros de la *Dark web*, donde algunas plataformas requerían invitaciones o procesos de verificación adicionales para permitir la consulta de contenido. Además, se detectó variabilidad en la disponibilidad de fuentes, ya que algunas filtraciones desaparecieron o cambiaron de ubicación antes de ser analizadas en profundidad, lo que generó la necesidad de diversificar los puntos de búsqueda.

Por último, durante el proceso de validación de datos, se identificaron algunos falsos positivos en filtraciones atribuidas a la empresa objetivo. Esto supuso un esfuerzo adicional en la verificación y cruce de información para garantizar la fiabilidad de los hallazgos.

A pesar de estos inconvenientes, todas las incidencias fueron gestionadas de manera efectiva, permitiendo la correcta ejecución del proyecto sin comprometer la calidad de los resultados obtenidos.

Autor:	Eduardo Blanco Bielsa		© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria			Hoja 28 de 81



## Riesgos

Riesgo	Descripción	Impacto	Medidas de mitigación	Eventos relacionados (bitácora de incidencias)
Bloqueo de cuentas	Algunas plataformas bloquean cuentas por actividad sospechosa o automatizada.	Medio	Uso de diferentes cuentas, <i>proxies</i> y modificación de patrones de acceso.	Se produjeron bloqueos temporales y permanentes en cuentas utilizadas para la investigación.
Dificultades en el <i>web scraping</i>	Implementación de <i>captchas</i> , restricciones <i>IP</i> y mecanismos <i>anti-bot</i> .	Medio	Uso de rotación de IP (mediante <i>VPNs</i> ), ajustes en el tiempo de consulta y técnicas de <i>web scraping</i> avanzadas.	Se encontraron dificultades en la extracción de datos de ciertos sitios protegidos.
Exposición a <i>malware</i>	Riesgo de infección por archivos maliciosos en foros de la <i>Dark web</i> .	Alto	Uso de entornos aislados, análisis de archivos antes de su apertura con múltiples herramientas y acceso controlado a sitios web sospechosos.	No se identificaron incidentes de infección, pero se tomaron precauciones estrictas.



Trazabilidad de la investigación	Posibilidad de que la actividad de recolección de datos sea detectada y registrada, así como identificación de la identidad del investigador.	Medio	Uso de redes privadas virtuales <i>VPNs</i> , navegadores con enfoque en privacidad y herramientas de anonimización.	No se han identificado intentos de trazabilidad, pero se mantuvo un monitoreo constante.
Falta de información pública	Usuarios con configuraciones de privacidad restrictivas o sin presencia en redes sociales.	Bajo	Ampliación de fuentes de búsqueda y exploración de foros alternativos.	En algunos casos la información fue limitada, reduciendo la efectividad del análisis <i>OSINT</i> .

*Tabla 3. Riesgos identificados.*

# Cierre del proyecto

## Planificación final

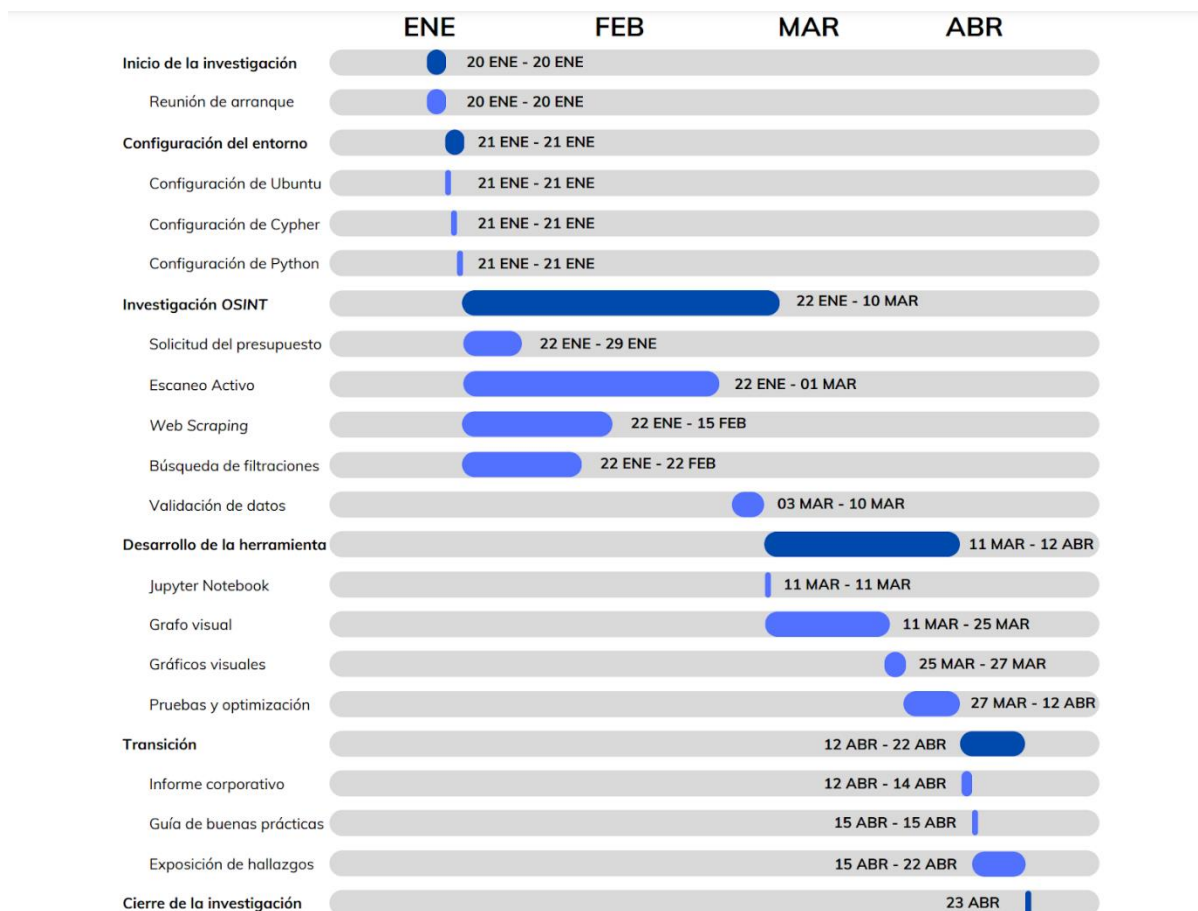


Ilustración 5. Planificación final.

Cabe destacar que, durante el transcurso del proyecto, se produjo un retraso de aproximadamente 10 días naturales respecto a la planificación inicial, debido a complicaciones técnicas surgidas en el desarrollo de la herramienta de visualización. Este desfase fue gestionado adecuadamente, permitiendo completar el proyecto dentro de los márgenes razonables y sin comprometer la calidad de los resultados obtenidos.





## Informe final de riesgos

Tras la finalización del proyecto, se realizó un análisis retrospectivo de los riesgos identificados y su impacto en la investigación. A continuación, se presenta un resumen de su ocurrencia y la efectividad de las medidas de mitigación implementadas.

1. **Bloqueo de cuentas:** ocurrió de forma parcial, afectando temporalmente algunas cuentas utilizadas para la recopilación de información. Sin embargo, la creación de cuentas alternativas y el uso de proxies permitió continuar con el proceso sin interrupciones significativas.
2. **Dificultades en el web scraping:** se presentaron desafíos en la extracción de datos debido a captchas y mecanismos de protección contra bots. El uso de técnicas avanzadas de *scraping* y la adaptación de tiempos de consulta mitigaron en gran medida estos inconvenientes.
3. **Exposición a malware:** no se detectaron infecciones ni accesos a archivos maliciosos. El uso de entornos aislados y herramientas de análisis preventivo resultó efectivo para minimizar este riesgo.
4. **Trazabilidad de la investigación:** no se encontraron indicios de rastreo o identificación de la actividad realizada. El uso de VPNs y técnicas de anonimización protegió la identidad del investigador y la seguridad del proceso.
5. **Falta de información pública:** este riesgo tuvo un impacto notable en algunos casos, ya que hubo individuos sin presencia en redes sociales o con configuraciones de privacidad restrictivas. A pesar de ello, la ampliación de fuentes de búsqueda permitió obtener información relevante en la mayoría de los casos.

## Presupuesto final de costes

Para la realización de esta investigación, se dispuso de un presupuesto inicial de **50,00 euros**, de los cuales se utilizaron **22,00 euros** en la adquisición de herramientas esenciales para el análisis OSINT.

El desglose del gasto final es el siguiente:

- **Licencia de la herramienta Lookup.io:** 12,01 euros.
- **Licencia de la herramienta Leakpeek.com:** 9,99 euros.

Estos pagos se realizaron el **24 de febrero de 2025**, empleando **Bitcoin** como método de pago debido a la naturaleza de las plataformas utilizadas y la necesidad de preservar

Autor:	Eduardo Blanco Bielsa			© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008	
Trabajo de Fin de Grado, Convocatoria Ordinaria				Hoja 32 de 81

la privacidad en las transacciones. En el momento de la compra, el valor de **1 BTC equivalía a 90.508,30 euros**, por lo que la cantidad exacta de Bitcoin empleada fue **0,000524 BTC**. Es importante destacar que, al realizar pagos en estas plataformas, se aplicó una comisión por transacción, habitual en este tipo de operaciones, que supuso un coste adicional aproximado de **0,0001 BTC (0,90 euros)**. Este importe se deriva del uso del monedero de las comisiones cobradas por el proveedor de la cartera digital (Binance) al realizar los pagos con Bitcoin a proveedores de herramientas OSINT.

Si bien el resto del presupuesto no fue empleado, es importante considerar que la investigación requirió recursos adicionales que fueron asumidos por el investigador, tales como:

- **Infraestructura computacional:** uso de un equipo portátil **Lenovo ThinkPad T14 Gen5 (AMD Ryzen 7 PRO)**, valorado en **1.479,00 euros**.
- **Conectividad a internet:** indispensable para la recopilación de datos y el acceso a foros especializados (**Movistar, 38 euros/mes x 3 meses = 114,00 euros**).
- **VPN profesional:** indispensable para la recopilación de datos y el anonimato del investigador (**NordVPN, 98,00 euros/año**).
- **Almacenamiento y copias de seguridad:** espacio necesario para la conservación de la información recopilada (**Disco SSD Samsung 990 Evo 1TB, 105,41 euros**).

Este desglose permite evidenciar una gestión transparente del presupuesto, asegurando la trazabilidad de los recursos empleados.

## Informe de lecciones aprendidas

Durante el desarrollo de esta investigación, se han adquirido conocimientos y habilidades clave en diversas áreas relacionadas con *OSINT*, Ingeniería social, ciberseguridad y *Big Data*. A continuación, se destacan las principales lecciones aprendidas.

- **Profundización en técnicas OSINT avanzadas:** si bien el conocimiento previo en *OSINT* era una base fundamental para la investigación, se explotaron nuevas técnicas de obtención y correlación de información, incluyendo:
  - **Anonimización y ocultación de trazabilidad:** Se implementaron estrategias para reducir la exposición digital del investigador, utilizando

Autor:	Eduardo Blanco Bielsa		© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria			Hoja 33 de 81

VPNs, proxies y puentes de la red Tor para acceder a ciertos recursos sin revelar la identidad ni ubicación del origen de la consulta.

- **Acceso a foros en múltiples idiomas:** Se investigaron comunidades de intercambio de información previamente desconocidas en inglés, ruso y chino, destacando cómo la barrera del idioma puede representar un obstáculo en la recopilación de datos y la verificación de su autenticidad. Se emplearon herramientas de traducción automatizada tales como **DeepL** y técnicas de análisis contextual para interpretar publicaciones en estos foros.
- **Monitoreo de grupos de cibercrimen y terrorismo:** Se identificaron y analizaron espacios donde operan colectivos como BelsenGroup y Lazarus, cuyos foros y canales de comunicación contienen filtraciones de datos sensibles y patrones de ataque dirigidos a empresas y gobiernos.
- **Verificación de la autenticidad de correos electrónicos:** Para garantizar la fiabilidad de los datos extraídos, se aplicaron técnicas como:
  - Análisis de registros SPF, DKIM y DMARC para determinar si un correo ha sido suplantado.
  - Análisis de archivos .EML para determinar direcciones de correo comprometidas.
  - Cruce de información en bases de datos de filtraciones para comprobar si una dirección ha sido previamente comprometida.
  - Investigación de metadatos en correos filtrados para identificar la fuente y estructura del mensaje original.
- **Uso de infraestructuras protegidas para el acceso a filtraciones:** Se emplearon métodos avanzados para acceder a bases de datos y leaks privados en foros de la *Dark web* sin comprometer la integridad del sistema utilizado, asegurando que la descarga y el análisis de credenciales expuestas se hicieran en un entorno aislado y seguro.
- **Desarrollo y optimización en Python:** el proceso de desarrollo de herramientas para la visualización y el análisis de datos permitió mejorar la eficiencia de la escritura y depuración de código Python. De esta forma, se profundizó en el conocimiento de las siguientes librerías:

Autor:	Eduardo Blanco Bielsa		© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria			Hoja 34 de 81

- **Neo4j:** utilizado para modelar y representar relaciones entre entidades (correos, perfiles sociales, filtraciones, ...) en una base de datos *Cypher* (base de datos NoSQL en grafo).
- **Plotly Express:** empleado para la creación de gráficos interactivos y visualización avanzada de datos obtenidos en la investigación. Su uso facilitó la interpretación de tendencias y patrones en la exposición de información.
- **Pyvis:** utilizado para la representación gráfica e interactiva de redes de relaciones, permitiendo visualizar las conexiones entre los datos analizados.
- **Uso de herramientas especializadas:** A lo largo del estudio, se emplearon diversas herramientas OSINT con ejecuciones en *CLI*, entornos web y *GUI*, lo que permitió combinar automatización, accesibilidad y visualización de datos. Aunque algunas eran de pago, se optimizó el uso de funcionalidades de prueba para obtener información clave sin exceder el presupuesto, maximizando así la eficacia de la investigación.
- **Limitaciones y obstáculos en la recolección de datos:** Se experimentaron dificultades con mecanismos de protección en plataformas analizadas, tales como *CAPTCHAs*, restricciones de acceso y bloqueos de cuentas. Estas incidencias no solo requirieron la adaptación de técnicas de recolección y estrategias para minimizar la detección, sino que también fortalecieron el pensamiento lateral y la capacidad de trabajar bajo presión. En varios momentos, la sensación de haber alcanzado un punto sin salida puso a prueba la resiliencia y la creatividad para encontrar nuevas vías de acceso a la información.
- **Exploración de nuevas comunidades y foros de hacking:** como se mencionó anteriormente, durante la investigación se identificaron y exploraron nuevos foros de la *Dark web* y *Deep web*. Este análisis permitió comprender mejor la dinámica de estos espacios, la comercialización de datos robados y la exposición de información corporativa en mercados ilícitos.
- **Concienciación sobre seguridad y privacidad:** El análisis de la exposición de datos corporativos y personales resaltó la importancia de la concienciación en ciberseguridad, tanto a nivel empresarial como individual. Se evidenció cómo pequeñas filtraciones pueden derivar en vulnerabilidades críticas, reforzando la

Autor:	Eduardo Blanco Bielsa		© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria			Hoja 35 de 81



necesidad de buenas prácticas en protección de datos y gestión de credenciales.

De esta forma, la investigación no solo ha permitido obtener resultados relevantes para la empresa Ebroker Insurance Technologies, S.A., sino que también ha supuesto una oportunidad de aprendizaje significativo en herramientas, técnicas y metodologías avanzadas de OSINT, así como denotar la realidad de la comercialización ilegal de los datos filtrados. Además, ha resultado impactante comprobar la cantidad de información personal expuesta en foros y mercados clandestinos, lo que evidencia la importancia de que tanto empresas como individuos monitoreen activamente las filtraciones de datos y adopten medidas para minimizar su nivel de exposición digital.

Autor:	Eduardo Blanco Bielsa		© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria			Hoja 36 de 81

## Capítulo 4. Estado actual de los conocimientos Científico-Técnicos

Para garantizar la relevancia y originalidad de la investigación, es fundamental analizar herramientas existentes en el ámbito *OSINT* que puedan aportar funcionalidades similares a la investigación y herramientas desarrolladas en este proyecto. Existen soluciones comerciales y de código abierto que agilizan el proceso de recolección y análisis, pero presentan limitaciones en cuanto a personalización, verificación manual y capacidad para realizar análisis avanzados sobre grandes volúmenes de datos.

En este capítulo se realiza un análisis comparativo entre la investigación llevada a cabo en este trabajo y herramientas ampliamente utilizadas en el ámbito *OSINT*, como Maltego, BlackBird y Spiderfoot. Se evaluarán sus capacidades, carencias y la forma en que este proyecto ha logrado suplir sus deficiencias mediante metodologías manuales y el uso del Big Data en una base de datos Cypher (Neo4j).

### Maltego

Maltego es una de las herramientas *OSINT* más conocidas y utilizadas en ciberseguridad e inteligencia, debido a su capacidad para automatizar la recolección de información y representar datos en un grafo de forma visual. Su funcionamiento se basa en módulos denominados “*transforms*”, que permiten extraer datos desde múltiples fuentes y establecer relaciones entre entidades como correos electrónicos, direcciones IP, dominios, redes sociales y filtraciones de datos.

Algunas de sus ventajas incluyen:

- **Automatización de la recolección de datos:** reduce el tiempo necesario para obtener información, lo que presenta una mayor comodidad.
- **Visualización en grafos:** representación gráfica intuitiva de las relaciones entre los datos.
- **Integración con bases de datos externas:** permite consultas en bases de datos *OSINT* ya existentes.

Sin embargo, a continuación, se exponen algunas de sus limitaciones en comparación con este proyecto:

Autor:	Eduardo Blanco Bielsa		© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria			Hoja 37 de 81

- **Dependencia de módulos predefinidos:** Maltego no permite flexibilidad en la selección de fuentes, ya que se basa en APIs externas que pueden estar restringidas en función de la geolocalización del investigador o ser de pago.
- **Falta de verificación manual:** los datos extraídos no pasan por un proceso de validación, lo que aumenta el riesgo de falsos positivos.
- **No permite análisis Big Data con Neo4j y Cypher:** la información recolectada se presenta en un grafo visual estático sin capacidad para realizar análisis avanzados con consultas personalizadas en Cypher.
- **Exportación limitada a formatos clásicos:** aunque permite exportación de datos, no soporta una exportación directa a bases de datos especializadas en grafos. Existe un *transform* que permite hacer una conversión a formato CSV y de ahí a *Cypher*, pero con una investigación grande y compleja el resultado requiere revisión manual, lo que hace muy incómoda la exportación de la información obtenida.

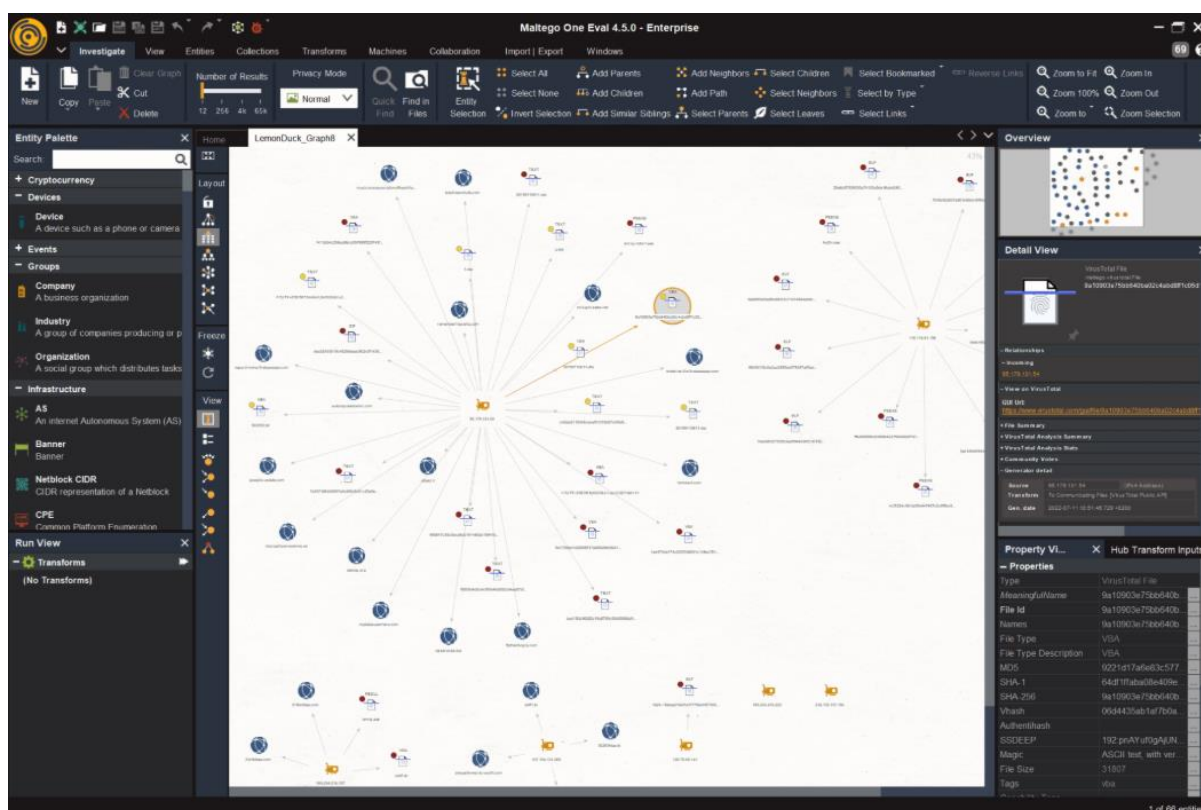


Ilustración 6. Maltego.

Autor:	Eduardo Blanco Bielsa	© 2025	
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria			Hoja 38 de 81



## BlackBird

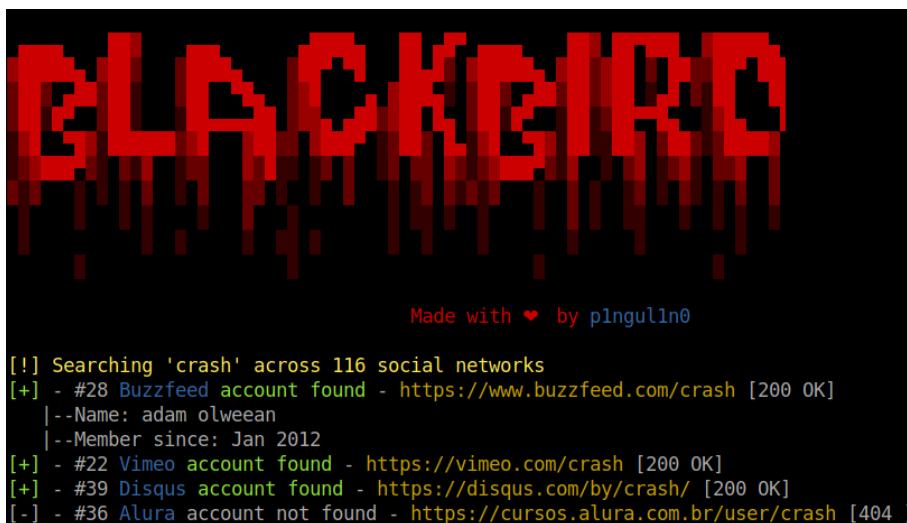
BlackBird es una herramienta especializada en recolección de información en redes sociales. Su función principal es identificar perfiles en diversas plataformas mediante usuarios, correos electrónicos o identificadores asociados.

Algunas de sus ventajas son:

- **Especialización en búsqueda de perfiles en redes sociales:** destaca por su capacidad de rastrear nombres de usuario, correos electrónicos y otros identificadores en diversas plataformas sociales. Esto permite identificar cuentas vinculadas a un individuo o empresa de forma rápida y eficiente.
- **Cruce automatizado de información en distintas plataformas:** la herramienta escanea diferentes plataformas simultáneamente, lo que facilita la detección de duplicados, alias o identidades falsas.

Sin embargo, a continuación, se exponen algunas de sus limitaciones en comparación con este proyecto:

- **No permite análisis avanzado con Big Data:** su enfoque se centra en la recopilación de datos, sin herramientas para análisis a gran escala.
- **No integra exportación a bases de datos orientadas a grafos:** solo permite exportaciones a formatos tradicionales como PDF o CSV.
- **No proporciona contexto sobre filtraciones:** no permite cruzar la información obtenida con bases de datos de *leaks*, lo que reduce la veracidad de los hallazgos.



```

      BLACKBIRD

      Made with ♥ by p1ngulln0

[!] Searching 'crash' across 116 social networks
[+] - #28 Buzzfeed account found - https://www.buzzfeed.com/crash [200 OK]
    |--Name: adam olweean
    |--Member since: Jan 2012
[+] - #22 Vimeo account found - https://vimeo.com/crash [200 OK]
[+] - #39 Disqus account found - https://disqus.com/by/crash/ [200 OK]
[-] - #36 Alura account not found - https://cursos.alura.com.br/user/crash [404 ]

```

Ilustración 7. BlackBird, CLI.

Autor:	Eduardo Blanco Bielsa		© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria			Hoja 39 de 81





## Spiderfoot

Spiderfoot es una herramienta *OSINT* de código abierto diseñada para automatizar la recopilación de información sobre direcciones IP, correos electrónicos, nombres de usuario, redes sociales y dominios.

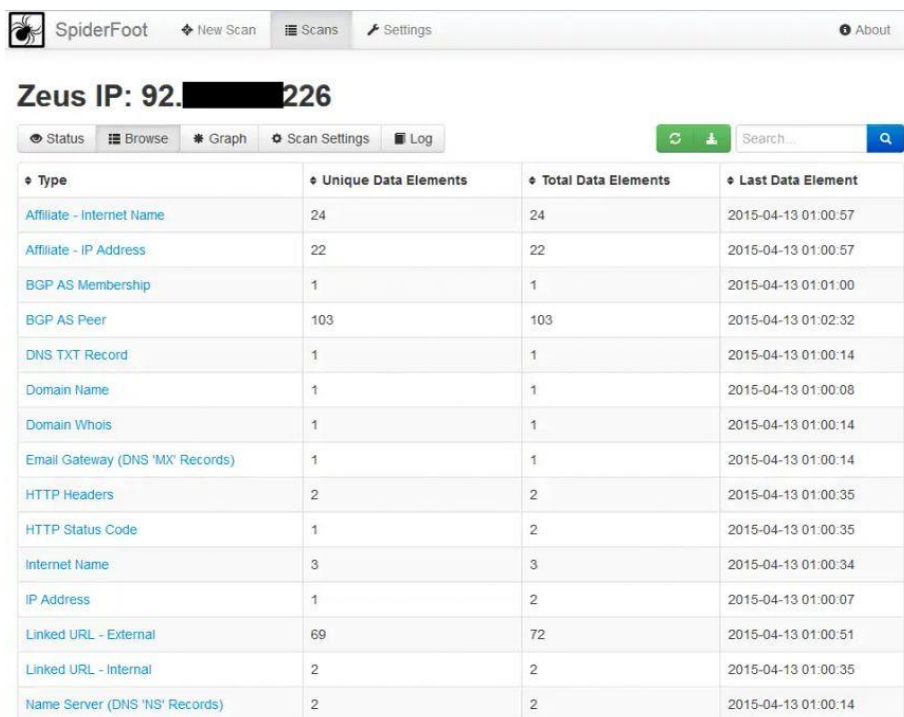
Algunas de sus ventajas son:

- **Automatización en la recopilación de datos desde múltiples fuentes:** destaca por su capacidad de extraer, correlacionar y analizar datos de manera automática, reduciendo el tiempo que el analista necesita para recopilar información manualmente.
- **Detección de posibles amenazas y vulnerabilidades asociadas:** además de recopilar datos, analiza patrones de riesgo en función de la información encontrada.
- **Capacidad de integración con bases de datos externas:** permite conectarse a diferentes *APIs* externas y bases de datos de ciberseguridad y filtraciones.

Sin embargo, a continuación, se exponen algunas de sus limitaciones en comparación con este proyecto:

- **Falta de personalización en la búsqueda de información:** aunque permite automatizar consultas, no proporciona la flexibilidad de un análisis manual detallado.
- **No permite modelado avanzado en grafos:** aunque almacena datos, no los organiza en una base de datos NoSQL como Neo4j.
- **Interfaz limitada en la visualización de datos complejos:** la forma de representar la información no permite el análisis visual en profundidad de conexiones entre elementos.
- **Exportación limitada:** únicamente permite exportar datos en formatos básicos como CSV y JSON, sin compatibilidad con exportación a bases de datos en grafo.

Autor:	Eduardo Blanco Bielsa		© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria			Hoja 40 de 81



SpiderFoot New Scan Scans Settings About

Zeus IP: 92.226.226.226

Status Browse Graph Scan Settings Log

Type	Unique Data Elements	Total Data Elements	Last Data Element
Affiliate - Internet Name	24	24	2015-04-13 01:00:57
Affiliate - IP Address	22	22	2015-04-13 01:00:57
BGP AS Membership	1	1	2015-04-13 01:01:00
BGP AS Peer	103	103	2015-04-13 01:02:32
DNS TXT Record	1	1	2015-04-13 01:00:14
Domain Name	1	1	2015-04-13 01:00:08
Domain Whois	1	1	2015-04-13 01:00:14
Email Gateway (DNS 'MX' Records)	1	1	2015-04-13 01:00:14
HTTP Headers	2	2	2015-04-13 01:00:35
HTTP Status Code	1	2	2015-04-13 01:00:35
Internet Name	3	3	2015-04-13 01:00:34
IP Address	1	2	2015-04-13 01:00:07
Linked URL - External	69	72	2015-04-13 01:00:51
Linked URL - Internal	2	2	2015-04-13 01:00:35
Name Server (DNS 'NS' Records)	2	2	2015-04-13 01:00:14

Ilustración 8. Spiderfoot.

## Diferencias clave y valor agregado de este proyecto

A diferencia de las herramientas mencionadas, este proyecto ha permitido superar sus principales limitaciones mediante:

- **Mayor flexibilidad en la selección de fuentes:** se han consultado foros de la *Dark web*, filtraciones de datos públicas y privadas y redes sociales sin restricciones ni limitaciones impuestas por APIs.
- **Verificación manual de la información obtenida:** se han aplicado criterios de validación y de contraste de información para eliminar falsos positivos.
- **Uso de base de datos NoSQL orientada a grafos (Neo4j + Cypher):** se ha modelado la información en un entorno donde es posible realizar consultas avanzadas para detectar patrones y hacer estudios *Big Data*.
- **Análisis Big Data:** en lugar de almacenar la información en archivos PDF, XSLX, CSV, JSON o texto plano, se ha estructurado en una base de datos altamente escalable y optimizada para relaciones complejas.
- **Mejora en la visualización y análisis:** el uso de librerías como Pyvis y Plotly ha permitido representar la información de manera dinámica, algo que Maltego, BlackBird y Spiderfoot no pueden hacer con su sistema de exportación estático.

Autor:	Eduardo Blanco Bielsa	© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.	Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria	Hoja 41 de 81	



De este modo, este proyecto ha demostrado que la combinación de *OSINT* manual con técnicas de *Big Data* permite suplir las limitaciones de las herramientas automatizadas. Aunque este enfoque requiere más tiempo en la fase de recopilación de información, proporciona un método más preciso, adaptable y escalable para la investigación de la exposición de la huella digital, asegurando un análisis más riguroso y contextualizado de los datos obtenidos.

Autor:	Eduardo Blanco Bielsa		© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria			Hoja 42 de 81



## Capítulo 5. Descripción del Sistema

Este proyecto no sólo aborda la investigación *OS/INT* de manera tradicional, sino que introduce el uso de técnicas de *Big Data* para mejorar el análisis y visualización de la información recopilada. A diferencia de herramientas convencionales, que presentan los datos de forma estática y limitada a exportaciones en formatos estándar (PDF, CSV o JSON), este sistema permite estructurar la información en una base de datos NoSQL en grafo (Cypher) y representarla dinámicamente mediante Python, ofreciendo un amplio abanico de posibilidades adaptadas a las necesidades específicas de cualquier equipo de análisis de datos.

La principal innovación de este enfoque radica en su capacidad para establecer relaciones de probabilidad, realizar análisis temporales y generar gráficos de exposición que permitan detectar patrones y riesgos en los datos analizados. Gracias a la estructura de la base de datos empleada, se pueden realizar consultas avanzadas para identificar tendencias y correlaciones entre filtraciones de datos, conexiones entre usuarios y posibles vulnerabilidades en la exposición de información personal o corporativa.

### Ejemplos de análisis posibles

A continuación, se indica una serie de ejemplos de uso de la investigación mediante la herramienta creada. Es implícito recalcar que todos los datos incluidos han sido completamente anonimizados para proteger la confidencialidad de la huella digital de Ebroker Insurance Technologies, S.A.

Autor:	Eduardo Blanco Bielsa			© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008	
Trabajo de Fin de Grado, Convocatoria Ordinaria				Hoja 43 de 81

## Detección de empleados con mayor exposición

Mediante el análisis de redes sociales y filtraciones, se puede identificar qué empleados de la empresa tienen más datos expuestos en fuentes abiertas y filtraciones, permitiendo establecer niveles de riesgo:

Mapa de Árbol de Empleados con Filtraciones de Datos (Total: 57 empleados con leaks, 174 leaks)

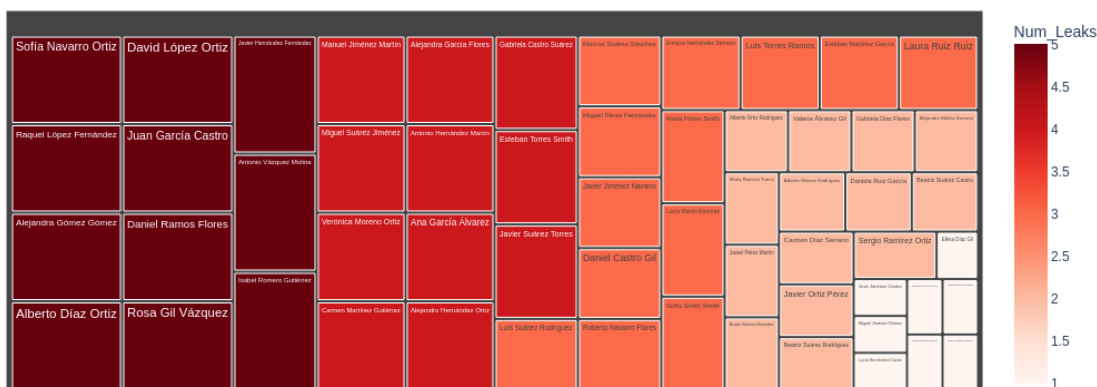


Ilustración 9. Detección de empleados con mayor exposición, formato treemap.

Empleados con Filtraciones de Datos (Total: 57 empleados con leaks, 174 leaks)

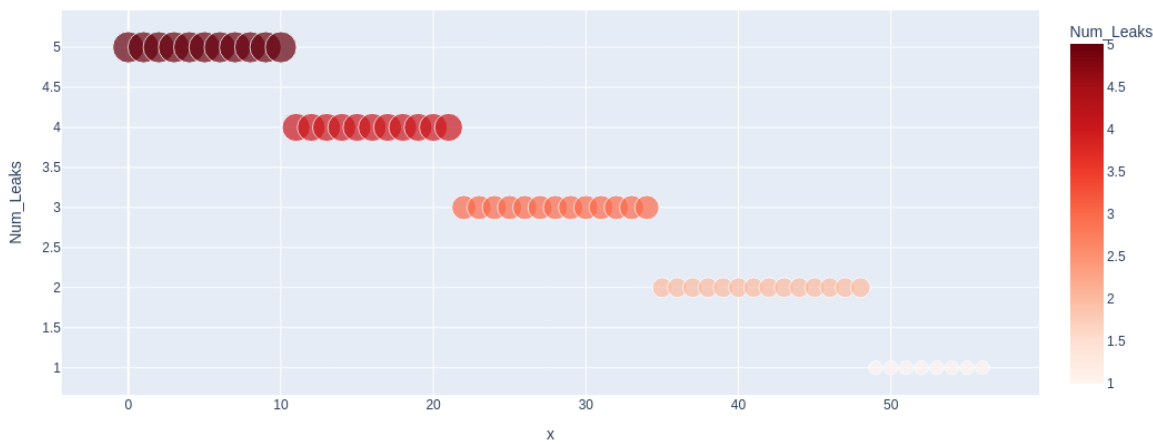


Ilustración 10. Detección de empleados con mayor exposición, formato scatter.

Este análisis permite a la empresa determinar qué trabajadores presentan un mayor riesgo de suplantación de identidad, *spear phishing* o ataques dirigidos, facilitando la adopción de medidas preventivas, como formación en ciberseguridad o modificación de credenciales.

## Relaciones entre empleados y sus perfiles sociales

Se analizó la presencia de los empleados en diversas plataformas sociales, identificando sus perfiles públicos y privados y su interconexión con la empresa. A través de este análisis se pueden identificar patrones de comportamiento, exposición de información sensible y relaciones interpersonales. Además, permite identificar información corporativa filtrada en redes sociales, lo que podría facilitar ataques de ingeniería social.

Distribución de Perfiles Sociales por Empleado



Ilustración 11. Relaciones entre empleados y sus perfiles sociales, formato treemap.

Este enfoque permite a la empresa identificar si la información personal de los empleados es utilizada en su contra o para generar un perfil detallado que podría ser explotado en ataques dirigidos, así como comprobar si dichos empleados exponen información corporativa o confidencial de forma pública.

## Plataformas sociales más frecuentes entre empleados

Mediante el análisis de las plataformas sociales más utilizadas por los empleados, se pudo determinar cuáles son las más vulnerables a la exposición de datos corporativos. Este análisis incluyó plataformas profesionales como LinkedIn, redes sociales de consumo general como Facebook o Instagram, y herramientas de comunicación como Disqus.

Número de Empleados por Plataforma Social (Total de Plataformas: 13)

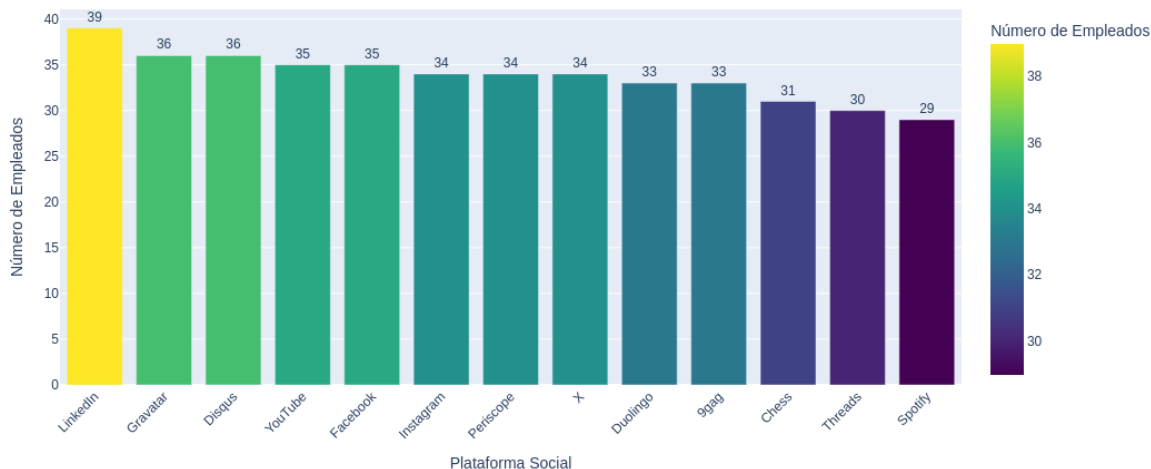


Ilustración 12. Plataformas sociales más frecuentes entre empleados, formato bar.

Número de Empleados por Plataforma Social (Total de Plataformas: 13)



Ilustración 13. Plataformas sociales más frecuentes entre empleados, formato treemap.

El conocimiento de las plataformas más frecuentes entre empleados ayuda a establecer políticas de seguridad más específicas y a desarrollar estrategias de formación basadas en los riesgos de cada plataforma.

Autor:	Eduardo Blanco Bielsa	© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo
Trabajo de Fin de Grado, Convocatoria Ordinaria		Versión: 2025.ES.008
		Hoja 46 de 81



# Capítulo 6. Metodología de trabajo

La metodología empleada en el presente trabajo se basa en un enfoque iterativo y estructurado, con el objetivo de garantizar la replicabilidad del entorno de análisis y maximizar la eficiencia de las herramientas desarrolladas. A continuación, se detalla cada una de las fases definidas previamente en el WBS.

## Creación del entorno de trabajo

El primer paso de la metodología consiste en la creación de un entorno controlado y replicable, diseñado para simular las condiciones reales de obtención y análisis de datos. Este entorno es fundamental para realizar los procesos de recopilación de información a partir de fuentes abiertas (*OSINT*), así como para validar la efectividad de las herramientas de análisis y visualización. La replicabilidad del entorno asegura que los resultados obtenidos puedan ser reproducidos en diferentes condiciones, lo que es esencial para la veracidad y la robustez de los resultados obtenidos.

A continuación, se indican las tareas llevadas a cabo en esta fase.

## Configuración de Ubuntu Desktop, Cypher y Jupyter Notebook

1. Descargar la última versión de Ubuntu disponible (en el caso de la investigación se corresponde con 24.04.2 LTS (Noble Numbat)). *También sería válido usar cualquier otra distribución de Linux u otro sistema operativo como Windows o MacOS, pero cambiarían los comandos de ejecución de Neo4j y Jupyter-Notebook.*
2. Creación de cuentas para la investigación (correos, carteras de criptomonedas, alias en la *Dark web* y cuentas de herramientas).
3. Descarga de Neo4j Desktop (habiendo configurado previamente la cuenta correspondiente). Ejecutar con el siguiente comando para abrir Neo4j Desktop (en mi caso la versión 1.6.1, sustituir x.y.z por la versión deseada y apuntar al directorio de instalación deseado):

```
./directorio/de/instalacion/neo4j-desktop-x.y.z-x86_64.AppImage --no-sandbox &
```

4. Crear un nuevo proyecto y base de datos en Neo4j. Abrir Neo4j Browser. Anotar las credenciales de acceso.
5. Instalar Python3 y crear un entorno virtual para la investigación (cambiar el directorio de instalación y el nombre de entorno por los deseados):

Autor:	Eduardo Blanco Bielsa		© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria			Hoja 47 de 81





```
sudo apt update && sudo apt install python3 python3-pip  
python3 -m venv /directorio/de/instalacion/NOMBRE_DEL_ENTORNO  
source /directorio/de/instalacion/NOMBRE_DEL_ENTORNO/bin/activate  
pip3 install notebook
```

6. Arrancar la Jupyter Notebook en el directorio de desarrollo de la investigación:

```
cd /directorio/de/investigacion && jupyter-notebook
```

## Investigación OSINT

En esta fase se realiza la recopilación y el análisis de datos mediante diversas técnicas y herramientas de OSINT. El proceso comienza con la solicitud de un presupuesto para establecer los recursos necesarios para poder llevar a cabo la investigación. Posterior y simultáneamente se realiza un escaneo activo, un web scraping y una búsqueda de filtraciones de la plantilla profesional de la empresa solicitante. Finalmente se realiza una validación y verificación de los resultados obtenidos, con el fin de asegurar la veracidad, validez y fiabilidad de la información antes de su uso por parte de la empresa.

### Solicitud del presupuesto

Esta etapa tiene como objetivo asegurar los recursos necesarios para la adquisición de licencias mínimas de determinadas herramientas especializadas que pueden ser utilizadas en el proceso de recopilación y análisis de datos. Dado que la investigación cuenta con una pequeña selección de herramientas de pago, es necesario contar con un presupuesto aprobado para proceder a la adquisición de dichas licencias.

Sin embargo, mientras se espera la aprobación de dicho presupuesto para la adquisición de las licencias, se puede comenzar el proceso de la investigación prescindiendo de dichas herramientas hasta una vez aprobado el presupuesto.

En caso de no ser aprobado, dichas herramientas no serán usadas en la investigación, lo que puede afectar a la efectividad de esta.

### Escaneo Activo, Web Scraping y Búsqueda de filtraciones

Para comprender mejor los conceptos asociados a esta fase, se indica a continuación en qué consiste cada etapa y los procedimientos aplicados en cada una:

Autor:	Eduardo Blanco Bielsa		© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria			Hoja 48 de 81



## Escaneo Activo

Hace referencia a la búsqueda manual de información en fuentes abiertas con el objetivo de identificar datos relevantes sobre la organización y su personal. Esta fase implica la consulta directa de sitios web corporativos (accesibles públicamente), publicaciones en medios de comunicación, blogs especializados y otros recursos accesibles públicamente que puedan contener información pertinente para la investigación.

Durante esta etapa se recopilan datos relacionados con la estructura organizativa, eventos relevantes, menciones en prensa y cualquier otro contenido que pueda contribuir al análisis de la exposición pública de la empresa.

A continuación, se indican los procedimientos empleados:

### *Análisis de la web corporativa*

- Se corresponde con el código [T1591](#) del *MITRE ATT&CK*.
- Revisión de secciones como “Quiénes somos”, “Equipo”, “Conócenos”, “Proyectos” y “Clientes” en busca de datos sobre empleados, departamentos y tecnologías utilizadas.
- Descarga y análisis de documentos accesibles en la web (PDF, DOCX, PPTX, ...) para extraer metadatos que puedan contener nombres de usuarios, rutas de archivos y versiones de software.

### *Búsqueda en medios de comunicación y blogs*

- Se corresponde con el código [T1593](#) del *MITRE ATT&CK*.
- Identificación de menciones sobre la empresa en medios digitales mediante consultas avanzadas en Google News y otras plataformas de noticias.
- Análisis de entrevistas y artículos donde se revelen datos sobre empleados, directivos o estrategias corporativas.

### *Uso de Google Dorks y operadores de búsqueda avanzada*

- Se corresponde con el código [T1593](#) del *MITRE ATT&CK*.
- Aplicación de los siguientes operadores para localizar información específica sobre la empresa en internet:

```
site:          intext:
intitle:       "<name>" (X OR Y OR Z)
inurl:         * (wildcards)
filetype: "doc | pdf | xls | txt | ps | rtf | odt | sxw | psw | ppt | pps | xml"
```

Autor:	Eduardo Blanco Bielsa	© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.	Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria	Hoja 49 de 81	



- Exploración de documentos indexados que puedan contener información confidencial.

#### *Investigación de perfiles públicos de la empresa (LinkedIn, Twitter, Spotify, Instagram...)*

- Se corresponde con el código [T1593](#) del MITRE ATT&CK.
- Obtención de información acerca de los empleados de la empresa, con sus respectivos puestos. De esta forma se puede contrastar la información obtenida en su web corporativa y ampliar la información personal de cada empleado de la plantilla.

#### *Identificación de correos corporativos*

- Se corresponde con el código [T1596](#) del MITRE ATT&CK.
- Se usan herramientas como Findymail, MailVerifier.io, Verifalia y Hunter.io para averiguar los correos electrónicos pertenecientes a la plantilla de la empresa.

### Web Scraping

Es una técnica que permite la extracción automatizada de información desde diversas fuentes digitales mediante el uso de herramientas específicas (códigos [T1589](#), [T1591](#), [T1593](#), [T1594](#) y [T1596](#) del MITRE ATT&CK), ya sean de línea de comandos, con interfaz gráfica (GUI) o a través de APIs.

Permite obtener grandes volúmenes de datos de manera eficiente y estructurada. Esto incluye la recopilación de perfiles en redes sociales, publicaciones en foros especializados y la extracción de metadatos de documentos públicos.

A continuación, se indican los procedimientos empleados:

#### *Uso de herramientas Open Source*

- Se emplean las siguientes herramientas para obtener correos, usuarios y perfiles sociales: BlackBird, OsintBuddy, Scylla, the Harvester, h8mail, reconsider, recon-ng, Shodan, Shodan-dorks, spiderfoot, linkook, Email-Username-OSINT, Maigret, Email Harvester, Watcher, Toutatis, holehe, osintgram, tookie-osint, bbot, sherlock, twint, phoneinfoga, Photon, Inoitsu y Ghost Track.

#### *Uso de herramientas de pago*

- Se emplean las siguientes herramientas (en su versión de prueba gratuita): IntelligenceX, LeackCheck, Pentester.com, Hunter.io y Breachsense.
- Se adquiere la licencia de las siguientes herramientas (dado su coste y sus resultados relevantes): Leakpeek.com y Leak.lookup.

Autor:	Eduardo Blanco Bielsa		© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria			Hoja 50 de 81



## Búsqueda de filtraciones

Tiene como finalidad la identificación y el análisis de bases de datos, *combos*, *wordlists*, *stealers* y *ransomwares* que puedan contener información sensible relacionada con la organización o sus empleados. Para ello, se recurre a plataformas especializadas y foros alojados en la *Dark web* donde se publican credenciales comprometidas, registros internos, documentos corporativos y otros tipos de información derivados de brechas de seguridad (código [T1597](#) del MITRE ATT&CK).

Esta fase implica el uso de herramientas de verificación y análisis de filtraciones, como plataformas de búsqueda de credenciales comprometidas, exploración en foros clandestinos a través de navegadores especializados y el estudio de bases de datos expuestas en ataques previos.

A continuación, se indican los procedimientos empleados:

### *Consulta en plataformas de bases de datos de filtraciones*

- Uso de las plataformas Have I Been Pwned, Dehashed, IntelligenceX, LeakCheck, Secureito, Inoitsu y Pentester.com.
- Pago de licencias de dos plataformas: Leakpeek.com y Leak.lookup.
- *Algunas de estas plataformas también se usan en la etapa anterior pues permiten filtrar tanto credenciales como correos de empleados.*

### *Exploración de foros en la Dark web*

- Acceso a mercados y foros de múltiples países donde se comercializan bases de datos filtradas, identificando información relevante sobre la organización. En concreto, los foros accedidos son Breachforums, Exploit.in, XSS.is, boo.wf, crackingx, Darkforums, PShack, nodo313, omniforums, cyberarsenal, leakzone, BlackHatWorld e in4.bz.
- Uso de navegadores como Tor o LibreWolf y motores de búsqueda como Ahmia, Brave, DuckDuckGo, DarkSearch, Haystak, Tor66 y NotEvil.

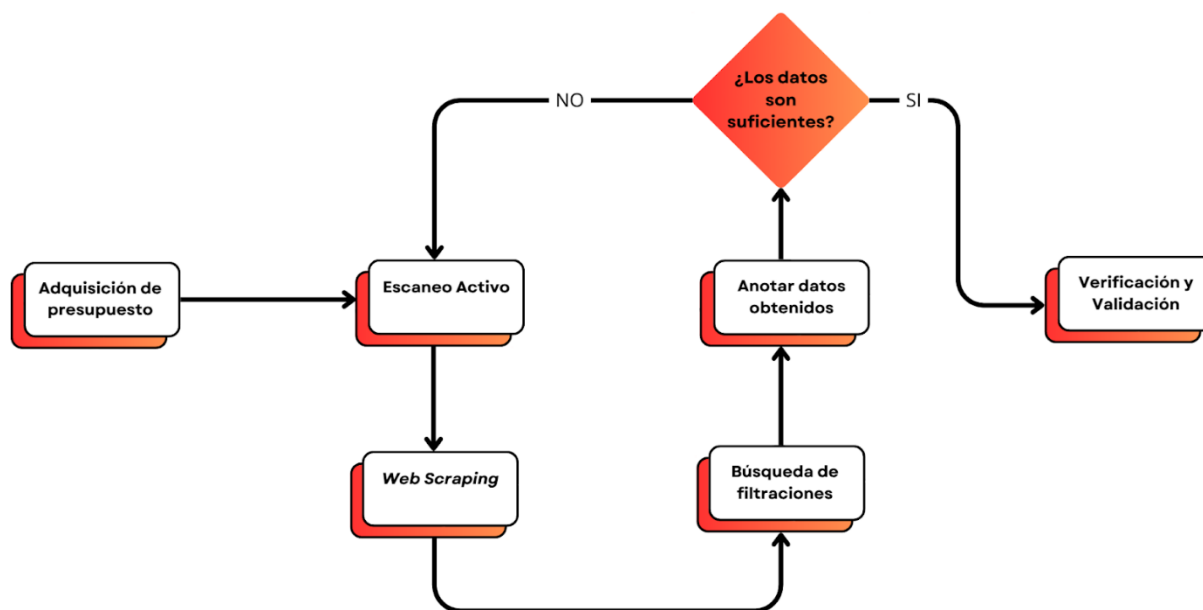
### *Adquisición, descarga y análisis de bases de datos, combos, wordlists e infostealers*

- Revisión de listas de combinaciones de correos electrónicos y contraseñas filtradas, identificando información relevante sobre la organización.
- Uso de herramienta de búsqueda de creación propia KeyHunter (cuyo desarrollo es ajeno a este proyecto) para indexar las bases de datos y buscar coincidencias relevantes para la organización.

Autor:	Eduardo Blanco Bielsa		© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria			Hoja 51 de 81

Es importante destacar que este proceso no es lineal, sino cíclico. El proceso debe repetirse tantas veces como sea necesario hasta obtener un volumen de información suficiente que permita realizar un análisis detallado sobre la exposición de la empresa solicitante. Cada nueva iteración puede proporcionar datos adicionales que refuercen los hallazgos previos o abran nuevas líneas de investigación.

A continuación, se incluye un diagrama de flujo indicando la metodología utilizada para realizar la fase de investigación:



*Ilustración 14. Diagrama de flujo de la metodología empleada en la investigación.*



# Capítulo 7. Resultados obtenidos

## Consideraciones previas

Debido a la naturaleza de esta investigación y al hecho de que los datos recopilados pertenecen a Ebroker Insurance Technologies, S.A., la información real obtenida no puede ser expuesta en este documento. Esto se debe a un acuerdo de confidencialidad, que protege la privacidad de los empleados y la seguridad de la empresa, evitando así cualquier riesgo derivado de la difusión de información sensible.

Para ilustrar los resultados obtenidos sin comprometer la integridad de los datos, se han generado investigaciones simuladas que reproducen fielmente los métodos y procesos empleados en el análisis real. Estas simulaciones permiten visualizar el potencial de la herramienta desarrollada y los alcances de la metodología aplicada, sin exponer información que pueda comprometer la seguridad de la empresa.

Más allá del caso específico de Ebroker Insurance Technologies, S.A., este estudio enfatiza la extrapolabilidad de la metodología empleada. La combinación de técnicas *OSINT* y *Big Data* aplicada en este proyecto puede ser utilizada por cualquier empresa que desee evaluar su nivel de exposición en fuentes abiertas. La flexibilidad del sistema desarrollado permite adaptarlo a múltiples escenarios, facilitando el análisis del impacto de la huella digital.

Esta versatilidad convierte al sistema en una herramienta adaptable y escalable, capaz de ajustarse a las necesidades de las distintas organizaciones para mejorar sus estrategias de ciberseguridad y protección de datos.

## Interpretación de los resultados

Para evaluar la efectividad de la metodología empleada y el sistema desarrollado, se ha generado una investigación ficticia basada en un conjunto de 70 empleados, replicando las condiciones de un caso real. En esta investigación, cada empleado tiene un número aleatorio de perfiles sociales (comprendido entre 0 y 9), un número aleatorio de brechas de datos (comprendido entre 0 y 3) y un número aleatorio de correos corporativos (comprendido entre 1 y 2 correos) de hasta dos compañías distintas (pues hay compañías que aún emplean dominios antiguos de correo).






El análisis se ha llevado a cabo empleando dos enfoques principales:

Autor:	Eduardo Blanco Bielsa			© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008	
Trabajo de Fin de Grado, Convocatoria Ordinaria				Hoja 53 de 81

## Interpretación mediante grafos

- Se ha representado la relación entre los empleados, sus perfiles en redes sociales y las filtraciones de datos asociadas utilizando una base de datos Cypher en Neo4j.
- Se puede identificar de un vistazo la jerarquía corporativa, así como los empleados con mayor número de filtraciones de datos y aquellos con las filtraciones más severas.
- Se han detectado patrones de interconexión entre empleados que comparten credenciales en diversas plataformas, lo que representa un riesgo potencial en caso de reutilización de contraseñas.

## Leyenda del grafo de la investigación

- Hay una imagen por cada empleado (en la investigación ficticia todos los empleados son generados por inteligencia artificial).
- El nodo naranja central se corresponde con el departamento al que pertenece uno o varios empleados: .
- Los iconos de redes sociales se corresponden con cada respectiva red social (X, Facebook, Instagram, Chess, etc.), así como los iconos de los correos privados (Gmail, Yahoo, etc). En la investigación ficticia al hacer clic en cualquier red social, se nos redirige a la web oficial de publicación de este trabajo, mientras que en una investigación real nos llevaría directamente a la red social del empleado en cuestión.
- En el caso de la investigación ficticia, que se ha creado con dos empresas, los siguientes iconos corresponden a los correos corporativos de la empresa ficticia uno y dos respectivamente:  y .
- Para las filtraciones de datos, hay tres categorías posibles:
  - **Low** (de riesgo bajo, ): corresponde a filtraciones de datos donde se expone cualquier tipo de información, salvo contraseñas de acceso, como DNIs, teléfonos, direcciones de vivienda, etc.
  - **Medium** (de riesgo medio, ): corresponde a filtraciones de datos donde se exponen contraseñas de acceso hasheadas o encriptadas,

Autor:	Eduardo Blanco Bielsa		© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria			Hoja 54 de 81

donde haría falta un servicio de fuerza bruta externo para descryptar o intentar averiguar dicha contraseña.

- **High** (de riesgo alto, 🔑): corresponde a filtraciones de datos donde se exponen contraseñas de acceso en texto plano, donde directamente cualquier atacante tiene acceso a las cuentas asociadas a dicha contraseña.

Para analizar el grafo de la investigación completa se ha usado la siguiente consulta Cypher:

```
query = "MATCH (n)-[r]->(m) RETURN n, r, m"
```

Lo que genera un HTML interactivo con todos los datos de la investigación de forma visual:

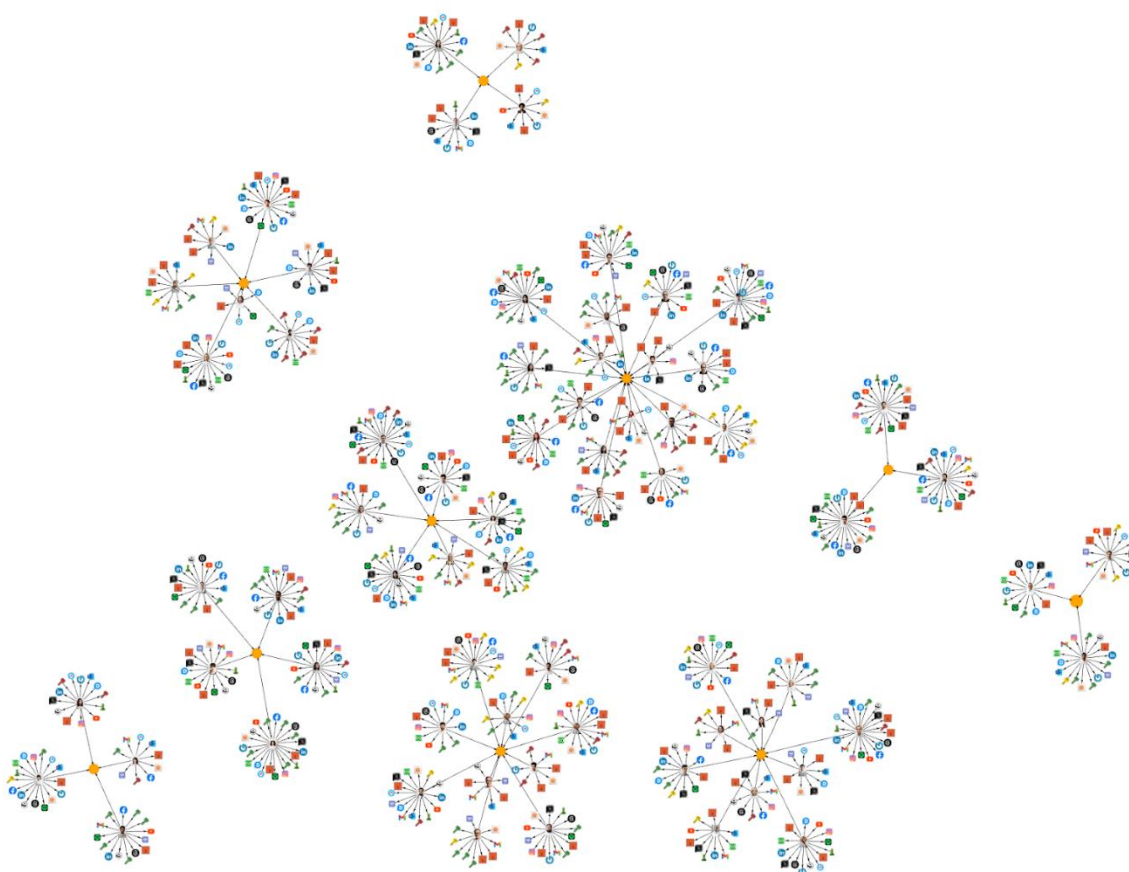


Ilustración 15. Grafo completo de la investigación.

Autor:	Eduardo Blanco Bielsa	© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo
Trabajo de Fin de Grado, Convocatoria Ordinaria		Versión: 2025.ES.008
		Hoja 55 de 81



Si ampliamos el grafo podremos verlo con más detalle:

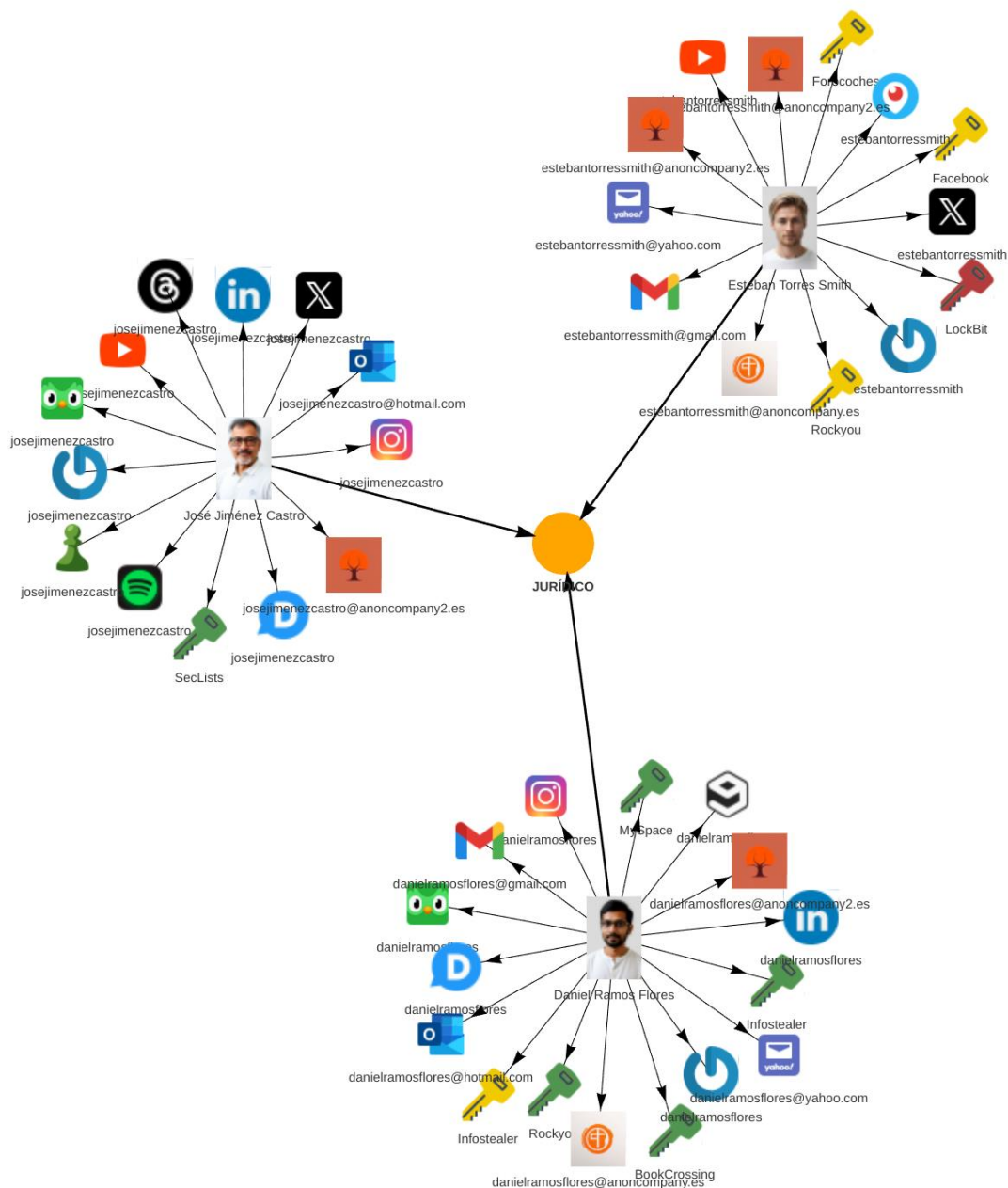


Ilustración 16. Parte ampliada del grafo completo de la investigación.

Autor:	Eduardo Blanco Bielsa	© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.	Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria		Hoja 56 de 81

Podemos emplear todo tipo de consultas para generar distintos tipos de grafos. A continuación, se incluyen ejemplos de consultas con los resultados obtenibles:

- Obtener empleados que trabajan en un área determinada.

```
MATCH (n:Empleado)-[r:TRABAJA_EN]->(m) RETURN n, r, m
```

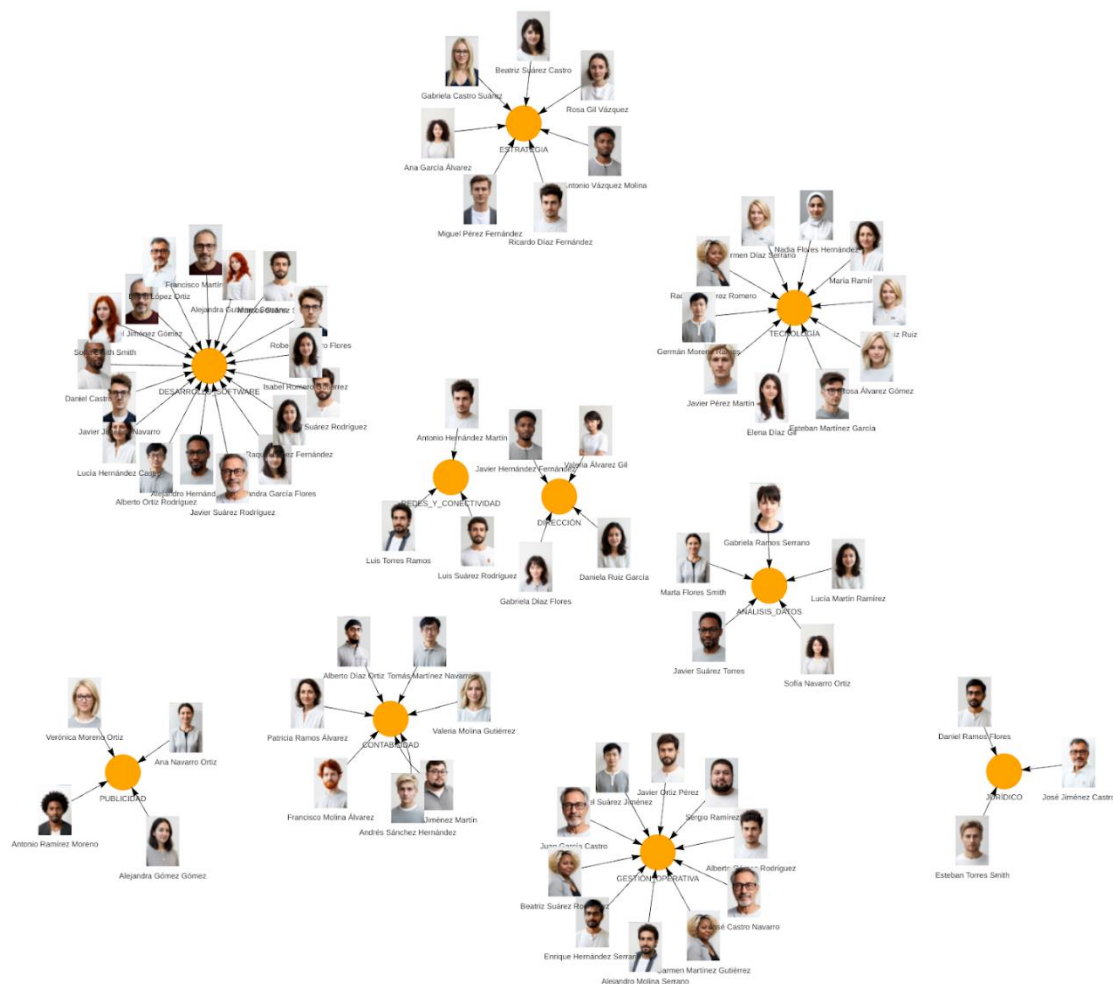
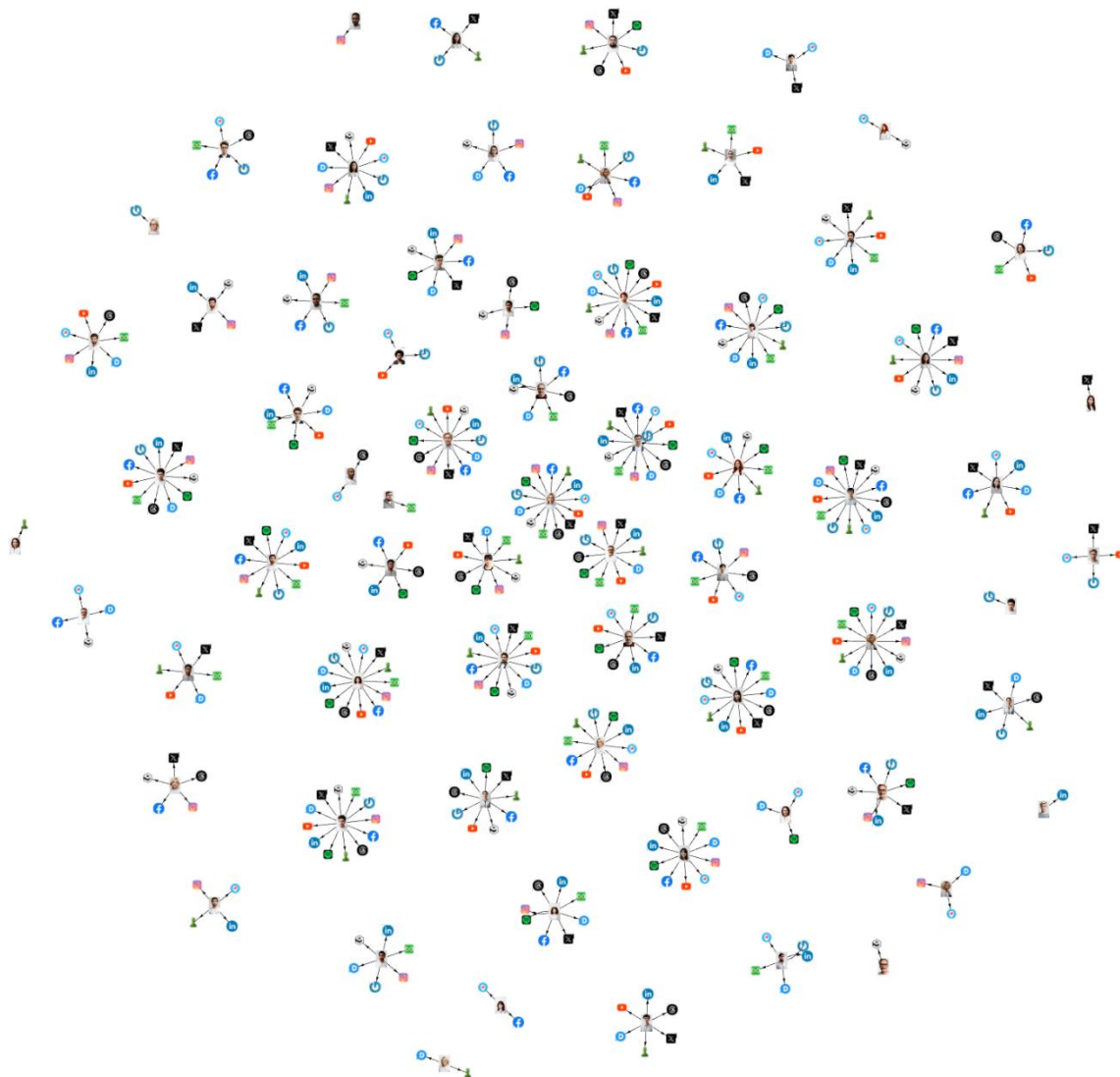


Ilustración 17. Grafo de los empleados con sus respectivos departamentos de trabajo.

- Obtener empleados que tengan uno o más perfiles sociales:

```
MATCH (n:Empleado)-[r:TIENE_PERFIL_EN]->(m) RETURN n, r, m
```



*Ilustración 18. Grafo de empleados junto con todos sus perfiles sociales.*

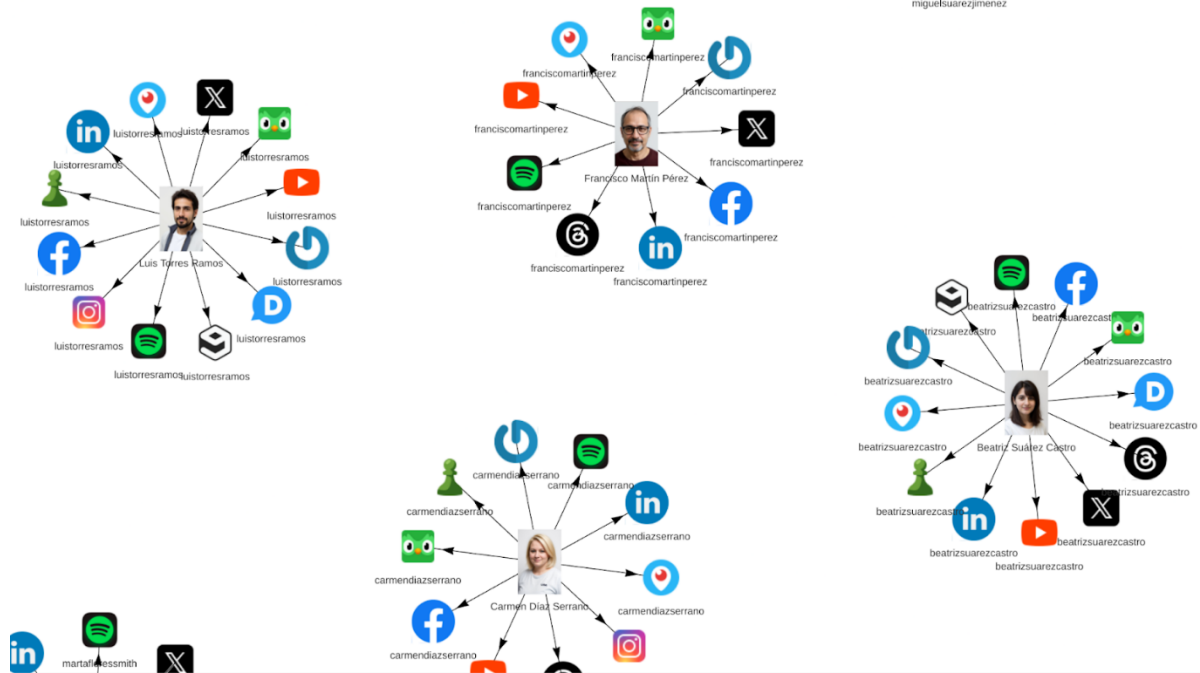
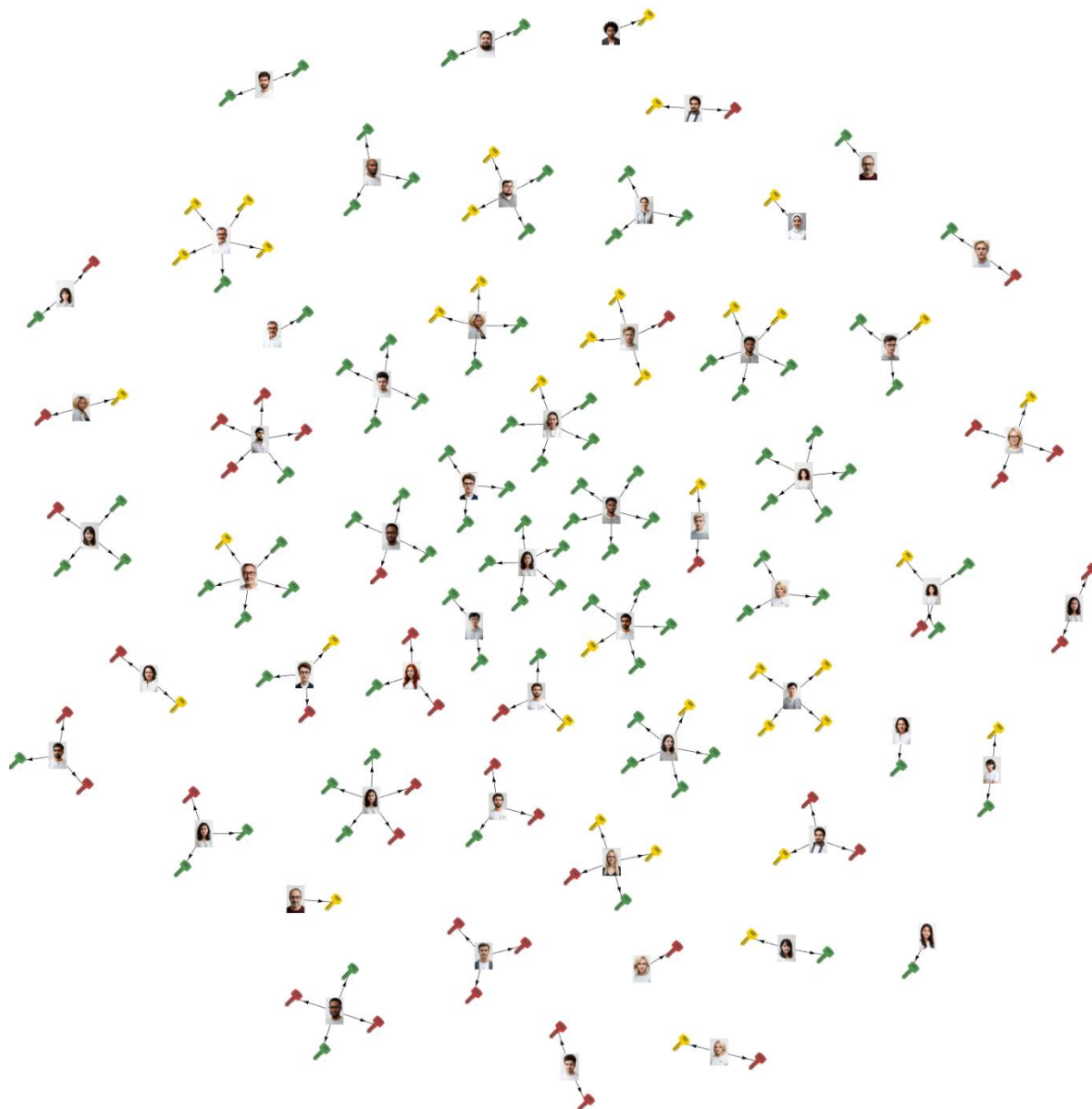


Ilustración 19. Parte ampliada del grafo de empleados junto con todos sus perfiles sociales.

Autor:	Eduardo Blanco Bielsa		© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria			Hoja 59 de 81

- Obtener empleados que tengan uno o más leaks:

```
MATCH (n:Empleado)-[r:TIENE_LEAK_EN]->(m) RETURN n, r, m
```



*Ilustración 20. Grafo de empleados con todas sus filtraciones de datos.*

Autor:	Eduardo Blanco Bielsa		© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria			Hoja 60 de 81

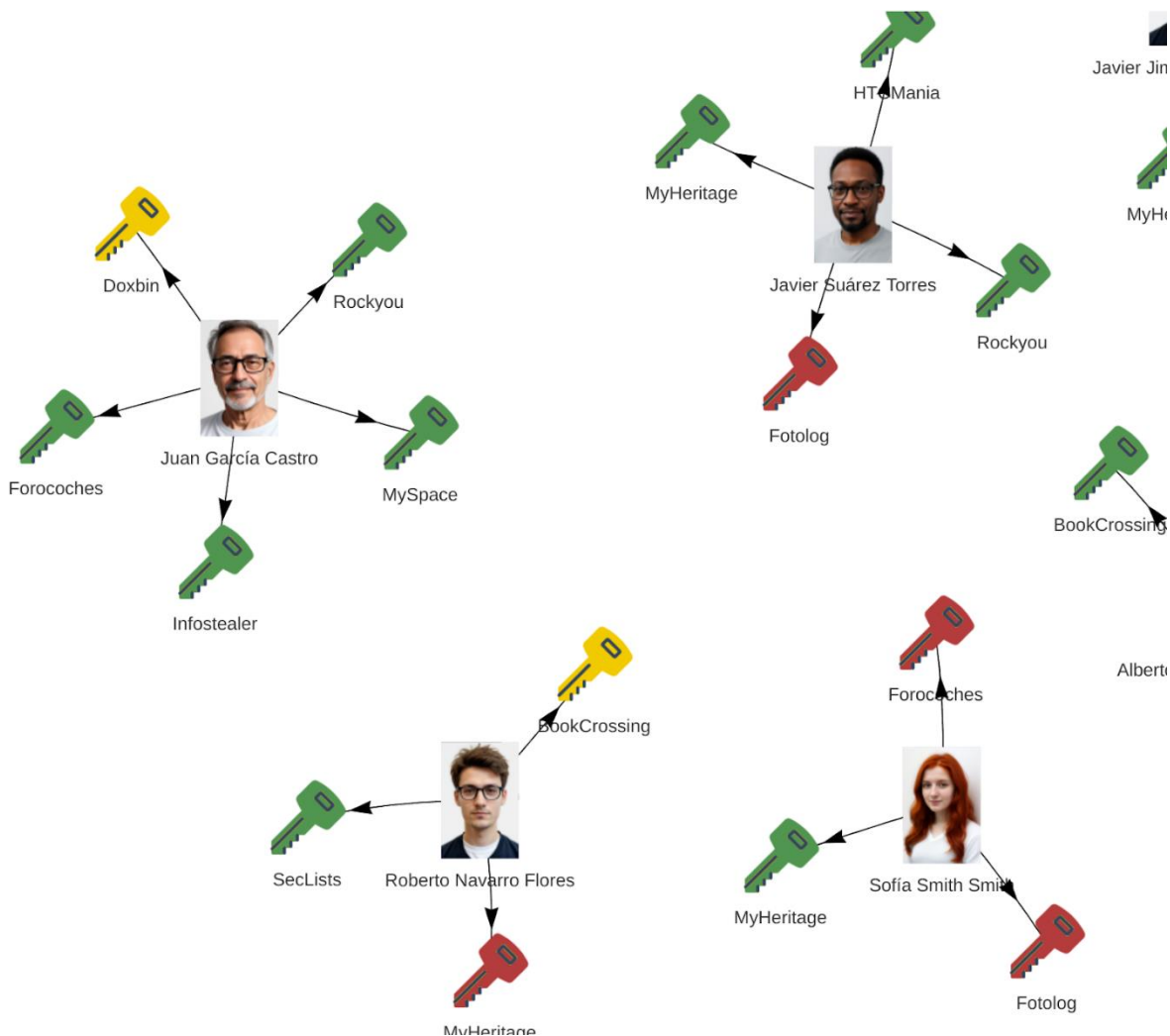


Ilustración 21. Parte ampliada del grafo de empleados con todas sus filtraciones de datos.

- Obtener todos los datos disponibles de un empleado (en este caso el empleado se llamará Esteban Torres Smith, generado por IA):

```
MATCH (n:Empleado)-[r]-(m) WHERE n.nombre = 'Esteban Torres Smith' RETURN n, r, m
```

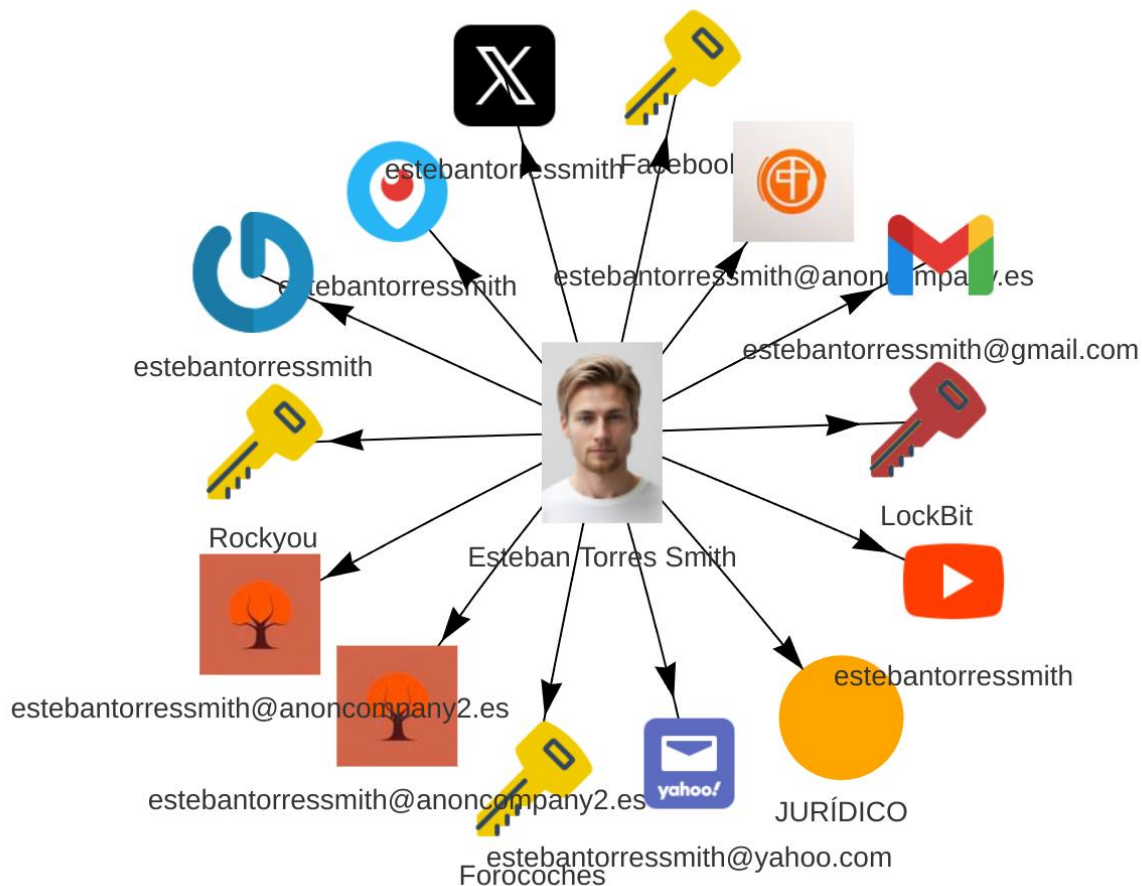


Ilustración 22. Grafo con todos los datos de la investigación sobre un empleado concreto.

Autor:	Eduardo Blanco Bielsa		© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria			Hoja 62 de 81



## Interpretación mediante *Big Data* y visualización de datos

Se han desarrollado cuatro ejemplos de análisis basados en técnicas *Big Data*, permitiendo visualizar patrones y correlaciones en los datos recopilados de múltiples formas. Cada uno de estos análisis se ha implementado a través de funciones específicas que generan gráficos interactivos para facilitar la interpretación de la información.

### Empleados agrupados por sus departamentos

- **Función:** `empleados_por_departamento_interactivo()`.
- **Objetivo:** Analizar la distribución de empleados según su departamento y evaluar qué áreas presentan un mayor nivel de exposición en función de su actividad digital.
- **Consulta Cypher:**

```
MATCH (e:Empleado) -[:TRABAJA_EN] -> (d:Departamento)
RETURN d.nombre AS Departamento, COUNT(e) AS Num_Empleados
```

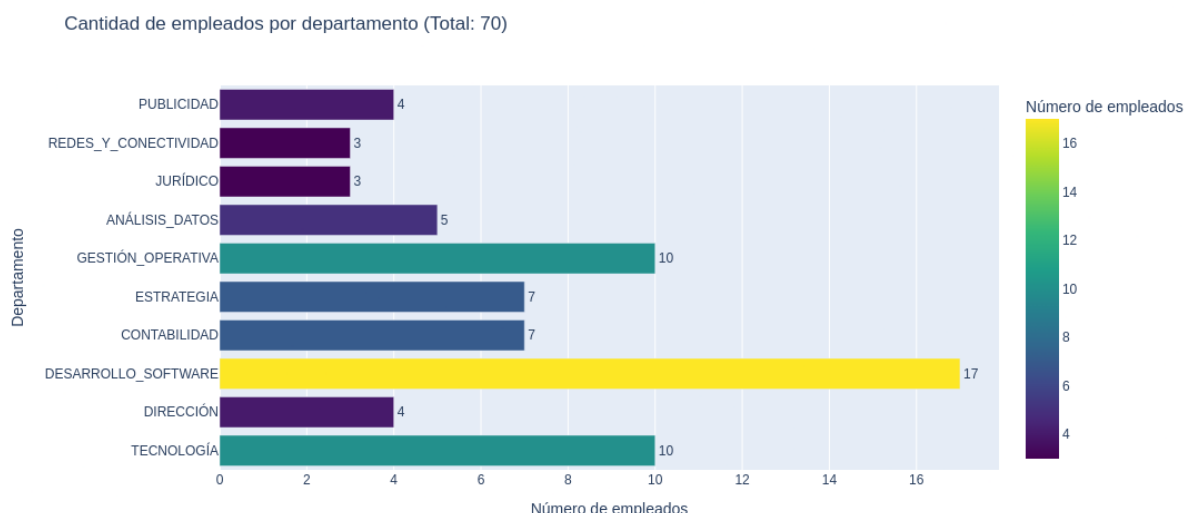


Ilustración 23. Gráfico de empleados agrupados por sus departamentos, formato bar.



## Plataformas más comunes donde los empleados tienen un perfil social

- **Función:** `empleados_por_plataforma()`.
- **Objetivo:** Determinar en qué plataformas sociales los empleados tienden a tener presencia digital, identificando aquellas más utilizadas y, por tanto, con mayor riesgo de exposición.
- **Consulta Cypher:**

```
MATCH (e:Empleado)-[:TIENE_PERFIL_EN]->(p:PerfilSocial)
RETURN p.plataforma AS Plataforma, COUNT(DISTINCT e) AS Num_Empleados ORDER BY
Num_Empleados DESC
```

Número de Empleados por Plataforma Social (Total de Plataformas: 13)

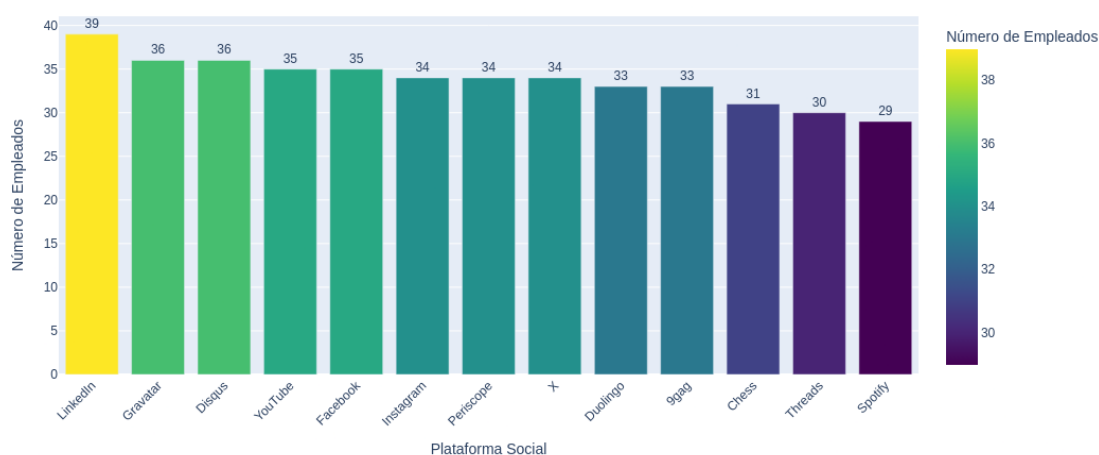


Ilustración 24. Gráfico de plataformas sociales más usadas por los empleados, formato bar.

Número de Empleados por Plataforma Social (Total de Plataformas: 13)

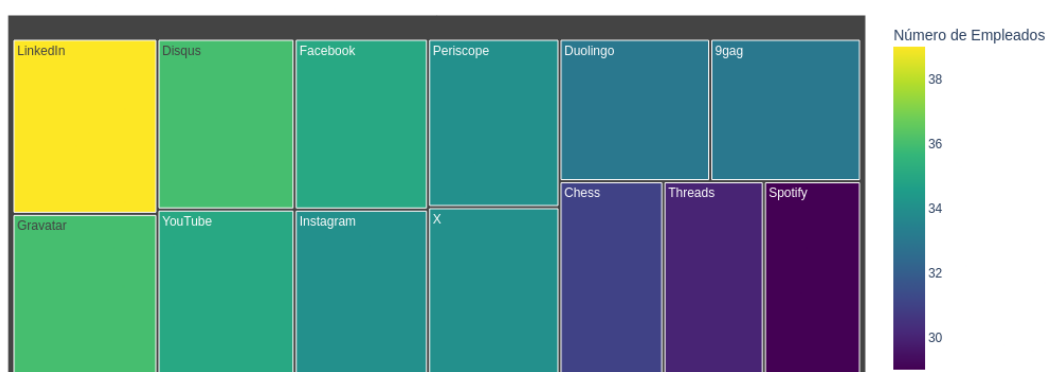


Ilustración 25. Gráfico de plataformas sociales más usadas por los empleados, formato treemap.

Autor:	Eduardo Blanco Bielsa	© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo
Trabajo de Fin de Grado, Convocatoria Ordinaria		Versión: 2025.ES.008
		Hoja 64 de 81



## Empleados con más brechas de datos

- **Función:** `empleados_con_leaks()`.
- **Objetivo:** Identificar a los empleados con más filtraciones de datos asociadas, permitiendo focalizar esfuerzos en concienciación y refuerzo de la seguridad.
- **Consulta Cypher:**

```
MATCH (e:Empleado)-[:TIENE_LEAK_EN]->(l:Leak)
RETURN e.nombre AS Empleado, COUNT(1) AS Num_Leaks ORDER BY Num_Leaks DESC
```

Empleados con Filtraciones de Datos (Total: 57 empleados con leaks, 174 leaks)

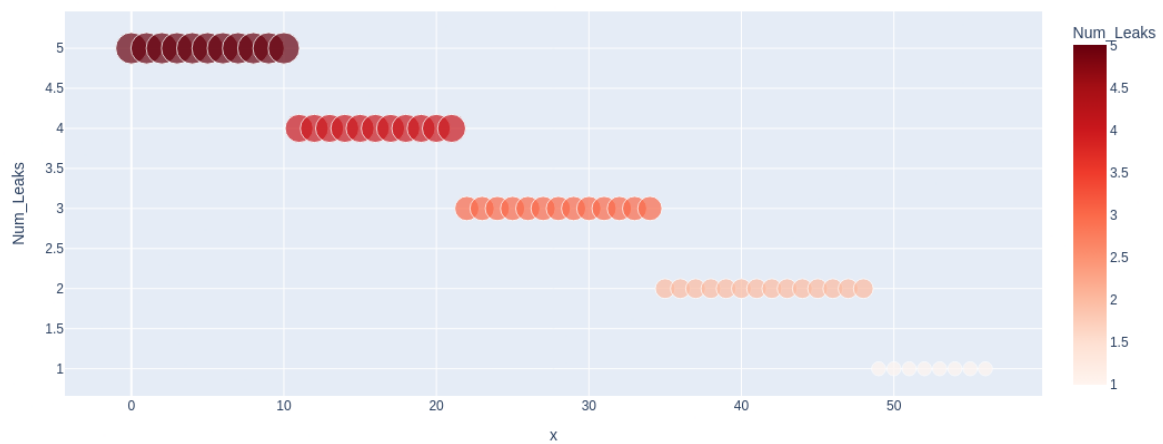


Ilustración 27. Gráfico de empleados con más brechas de datos, formato scatter.

Mapa de Árbol de Empleados con Filtraciones de Datos (Total: 57 empleados con leaks, 174 leaks)

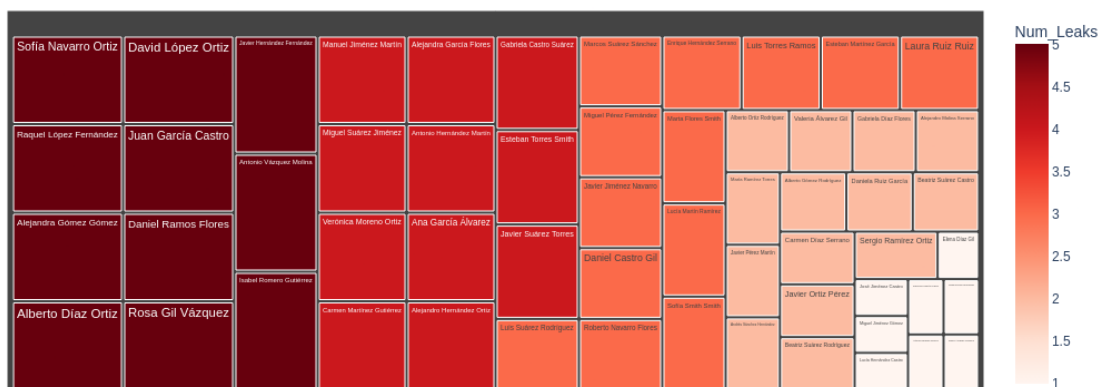


Ilustración 28. Gráfico de empleados con más brechas de datos, formato treemap.



Estos análisis permiten **representar gráficamente las tendencias y vulnerabilidades** detectadas en la investigación, facilitando la toma de decisiones estratégicas en materia de seguridad y privacidad.

## Discusión

La investigación realizada sobre la exposición de datos y las vulnerabilidades de Ebroker Insurance Technologies, S.A. ha sido crucial para detectar riesgos de seguridad y facilitar la toma de decisiones estratégicas en la empresa. Los hallazgos obtenidos permitieron a la organización no sólo identificar amenazas concretas, sino también establecer medidas correctivas y fortalecer su infraestructura de seguridad frente a posibles ataques futuros.

A continuación, se detallan los descubrimientos más relevantes y las acciones que se han tomado como consecuencia directa del estudio:

### Hallazgos clave y medidas adoptadas

#### Dispositivos comprometidos dentro de la red corporativa

Uno de los descubrimientos más críticos de la investigación fue la identificación de dos dispositivos dentro de la red interna de la empresa que habían sido comprometidos por *malware*:

- Un empleado tenía un **infostealer** activo en su equipo. Este tipo de malware está diseñado para robar credenciales, información personal y datos confidenciales almacenados en navegadores y aplicaciones. La presencia de este *software* malicioso representaba un riesgo crítico, ya que cualquier atacante con acceso a esas credenciales podría haber accedido a sistemas y servidores internos de la empresa.
- Otro empleado tenía un **troyano** en su sistema, lo que indicaba una posibilidad de acceso remoto no autorizado o la ejecución de acciones maliciosas en su dispositivo sin su consentimiento. Probablemente el *malware* permitía el acceso remoto desde un *Command and Control* (C2).

Se tomaron las siguientes medidas:

- Ambos dispositivos fueron aislados inmediatamente de la red corporativa para evitar cualquier posible propagación de amenazas.
- Se avisó a las personas afectadas y se les indicaron una serie de medidas a seguir para evitar que suceda lo mismo reiteradas veces.

Autor:	Eduardo Blanco Bielsa		© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria			Hoja 67 de 81



- Se reforzó la política de gestión de software, restringiendo la instalación de aplicaciones a ciertos grupos de empleados de la empresa.

## Implementación obligatoria de autenticación multifactor (MFA) en accesos críticos

La investigación reveló que varios empleados utilizaban contraseñas reutilizadas en múltiples plataformas, algunas de las cuales habían aparecido en bases de datos filtradas en foros de la *Dark web* o herramientas en línea. Este hallazgo evidenció la falta de seguridad en la gestión de las credenciales por parte de ciertos empleados de la empresa, lo que podría haber facilitado accesos no autorizados a sistemas internos de la misma.

Se tomaron las siguientes medidas:

- Se forzó el uso de autenticación multifactor (MFA) en todos los accesos a sistemas críticos, asegurando que incluso en caso de robo de credenciales, un atacante no pudiera acceder sin una segunda capa de verificación.

## Migración acelerada de controladores de dominio antiguos u obsoletos

Durante el análisis de los resultados de la investigación, se identificó que parte de la infraestructura de Ebroker Insurance Technologies, S.A. aún dependía de controladores de dominio antiguos, los cuáles aún no contaban con medidas de seguridad adecuadas. Estos sistemas representaban una vulnerabilidad significativa, ya que los atacantes podrían haber empleado algunas de las credenciales filtradas para obtener accesos o privilegios elevados en la red.

Se tomaron las siguientes medidas:

- Se decidió acelerar la migración de los controladores de dominio a versiones más recientes, con mejores estándares de seguridad y cifrado.

Se implementó una autenticación multifactor (MFA) para el acceso a dichos controladores.

Autor:	Eduardo Blanco Bielsa			© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008	
Trabajo de Fin de Grado, Convocatoria Ordinaria				Hoja 68 de 81



## Resumen final de resultados obtenidos

Hallazgo	Descripción	Medida adoptada
<b>Dispositivos comprometidos dentro de la red corporativa</b>	Se detectaron dos dispositivos infectados: uno con un <i>infostealer</i> y otro con un <i>troyano</i> . Riesgo crítico por posible robo de credenciales y acceso remoto no autorizado.	<ul style="list-style-type: none"><li>• Aislamiento inmediato de ambos dispositivos.</li><li>• Notificación a los empleados afectados con indicaciones de seguridad.</li><li>• Reforzamiento de políticas de instalación de software, restringiendo permisos a ciertos grupos.</li></ul>
<b>Falta de seguridad en la gestión de credenciales</b>	Se identificó uso de contraseñas reutilizadas por empleados, algunas de las cuales estaban comprometidas en filtraciones de la <i>Dark web</i> .	<ul style="list-style-type: none"><li>• Implementación obligatoria de autenticación multifactor (MFA) en accesos a sistemas críticos.</li></ul>
<b>Infraestructura con controladores de dominio obsoletos</b>	Parte de la infraestructura dependía de controladores de dominio antiguos sin medidas de seguridad adecuadas, lo que aumentaba el riesgo de escalamiento de privilegios mediante credenciales filtradas.	<ul style="list-style-type: none"><li>• Aceleración de la migración a controladores actualizados.</li><li>• Implementación de MFA para el acceso a los nuevos controladores.</li></ul>

Tabla 4. Resumen final de resultados obtenidos.



# Capítulo 8. Conclusiones y trabajo futuro

## Conclusiones

Este trabajo ha demostrado que el uso de *OSINT* combinado con técnicas de *Big Data* puede proporcionar una visión detallada y precisa sobre la exposición de datos personales y corporativos en fuentes abiertas. A lo largo de la investigación se ha logrado identificar vulnerabilidades reales en la empresa Ebroker Insurance Technologies, S.A., evidenciando la importancia de mantener un enfoque de seguridad proactivo y en constante evolución.

Los resultados obtenidos han confirmado que los métodos tradicionales de análisis *OSINT*, basados en herramientas comerciales automatizadas, no son tan claros ni suficientes para realizar investigaciones exhaustivas y adaptadas a cada caso particular. En contraste, la metodología empleada en este proyecto ha permitido:

- Correlacionar información dispersa en múltiples plataformas.
- Detectar amenazas, patrones de comportamiento y brechas de seguridad con mayor precisión.
- Estructurar y visualizar los datos de manera flexible mediante el uso de bases de datos NoSQL en grafo y técnicas avanzadas de análisis de datos.

Además, se ha cumplido con los objetivos propuestos inicialmente, destacando:

- La evaluación eficiente del nivel de exposición de la empresa.
- El análisis de los riesgos asociados a la disponibilidad de información sensible en el entorno digital.
- El uso de *Big Data* y técnicas *OSINT* de forma conjunta para obtener la mejor calidad de resultados y exposición de estos.

Asimismo, esta investigación ha servido para resaltar el **impacto real** de las filtraciones de datos, reforzando la importancia de mantenerse informado sobre brechas de seguridad y de adoptar medidas preventivas para minimizar los riesgos asociados.

Autor:	Eduardo Blanco Bielsa		© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria			Hoja 70 de 81



## Trabajo futuro

Esta investigación, que corresponde a la fase de **Reconocimiento** del marco *MITRE ATT&CK*, abre las puertas a la siguiente fase de **Explotación**, específicamente a la fase de **Acceso Inicial**. Esta fase podría llevarse a cabo mediante campañas de phishing o probando directamente las credenciales obtenidas en servidores y cuentas reales.

En este sentido, la empresa Ebroker Insurance Technologies, S.A. ha indicado que los datos obtenidos en esta investigación serán utilizados en futuras campañas de concienciación sobre phishing, como parte de un proceso formativo destinado a proteger a la empresa frente a esos ataques. Estas campañas estarán diseñadas para sensibilizar a los empleados acerca de los riesgos asociados al acceso no autorizado a información sensible y fortalecer las medidas de seguridad internas. Además, se pretende evaluar el nivel de vulnerabilidad de los empleados frente a ataques de ingeniería social, proporcionando retroalimentación inmediata y recomendaciones sobre cómo mejorar las prácticas de seguridad.

A través de este proceso, Ebroker tiene la intención de mejorar la resiliencia organizacional y optimizar sus defensas cibernéticas, reduciendo la probabilidad de que futuros incidentes de seguridad puedan tener un impacto significativo en la integridad de la empresa y sus recursos. Además, esta fase de explotación servirá como base para diseñar procedimientos y políticas más robustas para la protección de datos a nivel corporativo.

En cuanto al desarrollo de la herramienta utilizada en esta investigación, se considera que aún puede ser mejorada para ofrecer una solución más completa y personalizable. Esto incluiría la capacidad de generar los resultados obtenidos en diferentes formatos de salida en lugar de únicamente HTML, como informes PDF, archivos CSV o incluso dashboards interactivos mediante JavaScript, que permitan a los usuarios obtener un análisis más personal y accesible de los datos. También se podría crear una interfaz gráfica para hacerla más intuitiva y adaptable a las necesidades específicas de los usuarios, permitiendo una mayor flexibilidad en su uso en futuras investigaciones.

De cara al futuro, esta investigación podría extenderse a otras fases del *MITRE ATT&CK*, como Mantenimiento de Acceso y Exfiltración de Datos, para continuar ampliando el análisis de la exposición y diseñar estrategias defensivas aún más completas.

Autor:	Eduardo Blanco Bielsa			© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008	
Trabajo de Fin de Grado, Convocatoria Ordinaria			Hoja 71 de 81	





## Difusión de resultados

Se ha publicado en GitHub la herramienta que permite visualizar y analizar los datos, así como una base de datos de prueba → [Acceso a la herramienta y resultados de prueba](#).

## Capítulo 9. Bibliografía

- **Osint Framework**, “Colección de herramientas y recursos OSINT”, *Página web*, 2025, [[Enlace](#)].
- **MITRE ATT&CK**, “Marco de trabajo que recopila tácticas y técnicas basadas en observaciones de ciberataques reales”, *Página web*, 2025, [[Enlace](#)].
- **INCIBE**, “Instituto Nacional de Ciberseguridad Español”, *Página web*, 2025, [[Enlace](#)].
- **RGPD**, “Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos”, *Página web*, 2025, [[Enlace](#)].
- **Neo4j**, “Neo4j – Base de datos gráfica líder en el mundo”, *Página web*, 2025, [[Enlace](#)].
- **ElHacker.NET**, “Blog técnico de ciberseguridad”, *Página web*, 2025, [[Enlace](#)].
- **c1b3rn0t3s**, “Documentación técnica de ciberseguridad ofensiva”, *Página web*, 2025, [[Enlace](#)].
- **Tarlogic ciber inteligencia**, “Blog técnico sobre ciberinteligencia”, *Página web*, 2025, [[Enlace](#)].
- **Welivesecurity**, “Blog técnico de ciberseguridad”, *Página web*, 2025, [[Enlace](#)].
- **AwesomeOSINT**, “Repositorio técnico de ciberinteligencia”, *Repositorio web*, 2025, [[Enlace](#)].
- **OSINTforCountries**, “Repositorio técnico de ciberinteligencia”, *Repositorio web*, 2025, [[Enlace](#)].
- **fwhibbit.es**, “Blog técnico de ciberseguridad”, *Página web*, 2025, [[Enlace](#)].
- **Kaspersky blog**, “Blog técnico de ciberseguridad”, *Página web*, 2025, [[Enlace](#)].

Autor:	Eduardo Blanco Bielsa		© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria			Hoja 72 de 81



- **404 Media blog**, “Blog técnico (con algunos apartados de ciberseguridad)”, Página web, 2025, [[Enlace](#)].
- **Clandestine**, “Divulgador de información técnica de ciberseguridad”, Cuenta de X, 2025, [[Enlace](#)].
- **Phrack**, “Revista digital sobre ciberseguridad”, Página web, 2025, [[Enlace](#)].
- **Pyvis documentation**, “Documentación oficial de la librería de Python Pyvis”, Página web, 2025, [[Enlace](#)].

Autor:	Eduardo Blanco Bielsa		© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria			Hoja 73 de 81



# Capítulo 10. Anexos

## Plan de Gestión de riesgos

Riesgo	Descripción	Impacto	Respuesta
Bloqueo de cuentas	Algunas plataformas bloquean cuentas por actividad sospechosa o automatizada.	Medio	Se debe <b>reducir</b> este riesgo mediante el uso de diferentes cuentas, <i>proxies</i> y modificación de patrones de acceso.
Dificultades en el <i>web scraping</i>	Implementación de <i>captchas</i> , restricciones <i>IP</i> y mecanismos <i>anti-bot</i> .	Medio	Se debe <b>reducir</b> este riesgo mediante el uso de rotación de IP (mediante <i>VPNs</i> ), ajustes en el tiempo de consulta y técnicas de <i>web scraping</i> avanzadas.
Exposición a <i>malware</i>	Riesgo de infección por archivos maliciosos en foros de la <i>Dark web</i> .	Alto	Se debe <b>evitar</b> este riesgo mediante el uso de entornos aislados, análisis de archivos antes de su apertura con múltiples herramientas y acceso controlado a sitios web sospechosos.
Trazabilidad de la investigación	Posibilidad de que la actividad de recolección de datos sea detectada y registrada, así como identificación de la identidad del investigador.	Medio	Se debe <b>reducir</b> este riesgo mediante el uso de redes privadas virtuales <i>VPNs</i> , navegadores con enfoque en privacidad y herramientas de anonimización.



Falta de información pública	Usuarios con configuraciones de privacidad restrictivas o sin presencia en redes sociales.	Bajo	Se debe <b>aceptar</b> este riesgo si mediante la ampliación de fuentes de búsqueda y exploración de foros alternativos no se halla información relevante.
------------------------------	--	------	--

Tabla 5. Plan de Gestión de riesgos.



## Glosario de términos

- **Malware:** software malicioso diseñado para infiltrarse, dañar o explotar sistemas informáticos sin el consentimiento del usuario.
- **Ransomware:** tipo de malware que cifra los archivos del sistema de la víctima y exige un rescate económico, generalmente en criptomonedas, para restaurar el acceso a los datos.
- **Huella digital:** conjunto de datos que una persona u organización deja de forma voluntaria o involuntaria al interactuar en el entorno digital, como publicaciones en redes sociales o metadatos de navegación.
- **Ingeniería social:** técnicas de manipulación psicológica empleadas para engañar a las personas y obtener información confidencial, comúnmente utilizadas en ataques de phishing o fraudes.
- **Dark web:** parte de Internet que no está indexada por motores de búsqueda convencionales y que requiere software específico (como Tor) para acceder. Suele estar asociada a la privacidad extrema y, en algunos casos, a actividades ilícitas.
- **Deep web:** conjunto de contenidos web no indexados por los buscadores tradicionales. Incluye bases de datos privadas, correos electrónicos, intranets o foros protegidos con credenciales.
- **OSINT (Open Source Intelligence):** disciplina basada en la recolección y análisis de información pública y accesible libremente con el fin de obtener inteligencia útil, frecuentemente utilizada en ciberseguridad.
- **Captcha:** prueba automática diseñada para diferenciar entre humanos y bots. Suele utilizarse para prevenir accesos automatizados no autorizados a servicios web.
- **Bot:** programa automatizado diseñado para realizar tareas repetitivas sin intervención humana. Puede utilizarse tanto con fines legítimos como maliciosos.
- **Phishing:** técnica de fraude digital que consiste en suplantar la identidad de entidades legítimas para engañar a usuarios y obtener credenciales, datos personales o financieros.

Autor:	Eduardo Blanco Bielsa		© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria			Hoja 76 de 81



- **Bitcoin:** criptomoneda descentralizada basada en tecnología blockchain, utilizada frecuentemente como medio de pago en entornos donde se busca anonimato, incluida la Dark web.
- **VPN (Virtual Private Network):** tecnología que permite establecer una conexión segura y cifrada entre un dispositivo y una red, protegiendo la privacidad del usuario y ocultando su dirección IP.
- **Proxy:** servidor intermedio que actúa como puente entre un cliente y un servidor, utilizado para anonimizar conexiones o acceder a contenidos restringidos geográficamente.
- **Big Data:** conjunto de tecnologías, técnicas y procesos que permiten almacenar, procesar y analizar grandes volúmenes de datos, normalmente en tiempo real y con un enfoque predictivo.
- **Leak (o brecha de datos):** filtración de información, ya sea por una brecha de seguridad o divulgación no autorizada. Puede incluir contraseñas, correos electrónicos u otros datos sensibles.
- **CLI (Command Line Interface):** interfaz de usuario basada en texto que permite interactuar con un sistema operativo o programa mediante comandos escritos.
- **GUI (Graphical User Interface):** interfaz gráfica de usuario que permite interactuar con programas a través de elementos visuales como botones, menús o ventanas.
- **Spear phishing:** variante del phishing tradicional, más dirigida y personalizada, donde el atacante adapta su mensaje específicamente para una víctima concreta, aumentando su efectividad.
- **API (Application Programming Interface):** conjunto de definiciones y protocolos que permiten que diferentes aplicaciones se comuniquen entre sí. Facilitan el acceso a funcionalidades o datos de terceros.
- **Combo o wordlist:** colecciones de credenciales filtradas que combinan usuarios y contraseñas. Son utilizadas en ataques de fuerza bruta o verificación de accesos válidos en distintos servicios.
- **Stealer o infostealer:** tipo de malware diseñado específicamente para robar información sensible como contraseñas, cookies de navegador, wallets de criptomonedas, entre otros.

Autor:	Eduardo Blanco Bielsa		© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria			Hoja 77 de 81



- **Command and Control (C2):** infraestructura utilizada por atacantes para enviar órdenes a sistemas comprometidos y recibir información extraída de ellos. Es una parte fundamental del ciclo de vida del malware.

Autor:	Eduardo Blanco Bielsa		© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria			Hoja 78 de 81



## Herramientas utilizadas

- [Leakpeek](#).
- [Leak-Lookup](#).
- [Pryingdeep](#).
- [Blackbird](#).
- [OSINTBuddy](#).
- [Scylla](#).
- [theHarvester](#).
- [h8mail](#).
- [reconspider](#).
- [recon-ng](#).
- [spiderfoot](#).
- [linkook](#).
- [Email-Username-OSINT](#).
- [maigret](#).
- [EmailHarvester](#).
- [Watcher](#).
- [toutatis](#).
- [holehe](#).
- [Osintgram](#).
- [tookie-osint](#).
- [bbot](#).
- [hunter](#).
- [verifyemailaddress](#).
- [intelx](#).
- [email-rep](#).

Autor:	Eduardo Blanco Bielsa		© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria			Hoja 79 de 81





- [dehashed.](#)
- [KeyHunter.](#)
- [snusbase.](#)
- [leakcheck.](#)
- [darkint.](#)
- [sherlock.](#)
- [Snov.](#)
- [Verifalia.](#)
- [findymail.](#)
- [mailverifier.](#)
- [Shodan.](#)
- [linkook.](#)
- [Watcher](#)
- [toutatis.](#)
- [twint.](#)
- [phoneinfoga.](#)
- [Photon.](#)
- [inoitsu.](#)
- [GhosTrack.](#)
- [Leakcheck.](#)
- [Pentester.](#)
- [Breachsense.](#)
- [Have I Been Pwned.](#)
- [Dehashed.](#)
- [Secureito.](#)
- [Perchance AI.](#)

Autor:	Eduardo Blanco Bielsa		© 2025
Escuela de Ingeniería Informática, Ebroker Insurance Technologies, S.A.		Universidad de Oviedo	Versión: 2025.ES.008
Trabajo de Fin de Grado, Convocatoria Ordinaria			Hoja 80 de 81



ATT&CK<sup>®</sup>



Universidad de Oviedo