OSIRIS-GOV: Inteligência para Infraestruturas Resilientes e Segurança em Serviços Governamentais

Manuel Ramos Leite Carvalho Neto
Mestrado em Segurança Informática
Departamento de Ciência de Computadores
Faculdade de Ciências da Universidade do Porto
2025/2026

OSIRIS-GOV: Inteligência para Infraestruturas Resilientes e Segurança em Serviços Governamentais

Manuel Ramos Leite Carvalho Neto

Dissertação realizada no âmbito do Mestrado em Segurança Informática

Departamento de Ciência de Computadores

2025/2026

Orientador
Doutora Ivone de Fátima da Cruz Amorim, Professor Auxiliar Convidado, Faculdade de Ciências da Universidade do Porto

Coorientador
Doutora Eva Catarina Gomes Maia, Investigador Auxiliar, Instituto Superior de Engenharia do Porto

Supervisor externo na Entidade de Acolhimento
Pedro Norton Lages de Oliveira Barbosa, Founder, Skill & Reach

Agradecimentos

Embora a capa desta dissertação só contenha um nome, o trabalho nela vertido não teria sido possível sem o contributo de diversos outros, a quem aqui agradeço.

Em primeiro lugar, às minhas orientadoras, Professoras Ivone Amorim e Eva Maia, por terem aceitado o tema proposto num dos últimos dias possíveis para o efeito e por me terem acompanhado ao longo de toda esta longa jornada.

Em segundo lugar, ao meu supervisor, Pedro Barbosa, por, há já cerca de dois anos, me ter dado a oportunidade de fazer parte da Skill & Reach e, mais recentemente, da Skill & Research, assim como pela constante valorização do meu trabalho.

Em terceiro lugar, à minha família. Aos meus pais, por terem feito de mim aquilo que sou hoje e por tentarem sempre tirar todas e quaisquer pedras do meu caminho. Aos meus avós, Ana, Artur, Bebé e Né, não só pela construção da minha família, mas também pela capacidade de escrita, pela vocação para as tecnologias, pelos almoços e pela paciência, respetivamente. Aos restantes 31 membros da família, incluindo tios e primos, especialmente os mais novos, Kiko, Pia e Matilde, por toda a orientação.

Também à equipa – na verdadeira aceção da palavra – da Skill & Reach, por tanto. À Patrícia, pelos 64 bons e interessantes momentos de partilha, amizade, tolerância e incentivo/otimismo. À Sofia, por ser a alma e a diversão do escritório, por nos dar sempre motivos para (sor)rir e por nunca deixar de nos ouvir. Ao Tiago, pela pertinência dos nossos pontos de situação sobre cibersegurança, claramente visível na frequência com que ocorriam, quase todos os dias. Ao Vasco, por ver sempre todo o espectro de possíveis problemas que podem surgir, por mostrar que nem sempre tudo é preto no branco e pelo bom humor, boa disposição e clareza de todos os momentos.

Por último, aos meus amigos. À Biti, por me abrir as portas da Skill & Reach e de tudo o que se seguiu. À Carlota, por ser a amizade mais antiga e partilhar comigo um amor comum: o Futebol Clube do Porto. À Francisca, pela alegria constante e por estar sempre à altura das circunstâncias. À Isabel, por partilhar o mesmo amor pelo Porto e por ser a melhor companhia na neve e nos jogos no Dragão. À Maria, por estar sempre presente para me ouvir, aturar e suportar e por ser quase família. Ao Mini, aos Miguéis e ao Ricardo, por não se incomodarem por não os destacar individualmente.

A quem, por lapso, cansaço ou má memória, me possa ter esquecido de referir.

A todos, o meu muito obrigado. Esta é, sem dúvida, a página mais importante desta dissertação.

Resumo

Atualmente, a crescente digitalização da sociedade leva a que as organizações tendam a aumentar a sua dependência relativamente a sistemas e serviços de terceiros, expondo-se a um panorama de ciberameaças cada vez mais complexo. Este desafio é particularmente relevante no setor da administração pública, por ser um dos principais alvos dos ciberataques atuais. Neste contexto, a avaliação do ciber-risco organizacional torna-se essencial, não só por força dos riscos emergentes, mas também devido à necessidade de garantir a conformidade regulamentar. Embora se reconheça que esta análise deva abranger não só a própria organização, mas também os seus terceiros e prestadores de serviços, isto continua ausente na maioria das abordagens atuais.

Esta dissertação vai ao encontro deste problema ao apresentar o OSIRIS-GOV, enquanto solução de avaliação do ciber-risco organizacional. Para tal, o trabalho parte de uma revisão sistemática da literatura para levantamento do estado da arte, através da qual são comparadas as soluções de avaliação do ciber-risco organizacional e de avaliação do risco de terceiros. Com isto, são identificadas as suas principais limitações, como o facto de não considerarem o risco de terceiros no ciber-risco organizacional, a dificuldade no mapeamento da cadeia de fornecimento e a forte incidência sobre as vulnerabilidades nos ativos expostos à *Internet* para aferição do ciber-risco, em vez de outros fatores mais associados aos ciberincidentes atuais. O OSIRIS-GOV visa suprir estas limitações, consultando exclusivamente informação pública para mapear a cadeia de fornecimento de cada organização e considerando a exposição em fugas de dados das credenciais de cada entidade na avaliação do ciber-risco, com recurso a um classificador baseado em aprendizagem automática supervisionada.

Para validar a sua aplicabilidade, o OSIRIS-GOV foi aplicado ao caso de estudo da administração pública local portuguesa, mapeando a cadeia de fornecimento de 24 câmaras municipais e prevendo a probabilidade de ocorrência de um ciberincidente em 2026. Os resultados mostraram que as câmaras municipais de maior dimensão estão sujeitas a um maior ciber-risco, enquanto as câmaras municipais de menor dimensão têm, em proporção, um maior contributo dos seus terceiros para o respetivo ciber-risco.

Palavras-chave: Administração Pública, Avaliação do Ciber-Risco, Informação Pública, *Open-Source Intelligence*, Risco da Cadeia de Fornecimento, Risco de Terceiros

Abstract

Currently, the growing digitalisation of society means that organisations are becoming more reliant on third-party systems and services, exposing themselves to an increasingly complex cyber threat landscape. This challenge is particularly relevant in the public administration sector, as it is one of the main targets of current cyberattacks. In this context, cyber risk assessment becomes essential, not only because of emerging risks, but also due to the need to ensure regulatory compliance. Although it is recognised that this analysis should cover not only the organisation itself, but also its third parties and service providers, this remains absent from most current approaches.

This dissertation addresses this problem by presenting OSIRIS-GOV, a solution for organisational cyber risk assessment. To this end, the work begins with a systematic literature review to establish the state of the art, through which organisational cyber risk assessment and third-party risk assessment solutions are compared. This identifies their main limitations, such as the failure to consider third-party risk within organisational cyber risk, the difficulty in supply chain mapping, and the heavy focus on vulnerabilities in Internet-exposed assets for determining cyber risk, rather than on other factors more closely associated with current cyber incidents. OSIRIS-GOV aims to address these limitations by consulting exclusively public information to map each organisation's supply chain and by considering the exposure of each entity's credentials in data leaks when assessing cyber risk, using a classifier based on supervised machine learning.

To validate its applicability, OSIRIS-GOV was applied to the case study of Portuguese local public administration, mapping the supply chain of 24 town councils and predicting the probability of a cyber incident occurring in 2026. The results showed that larger town councils are subject to a greater cyber risk, whilst smaller town councils have, proportionally, a greater contribution from their third parties to their respective cyber risk.

Keywords: Cyber Risk Assessment, Open-Source Intelligence, Public Administration, Public Information, Supply Chain Risk, Third-Party Risk

Índice

Lista de Tabelas

Lista de Figuras

Lista de Abreviaturas

ACM	ASSOCIATION FOR COMPUTING MACHINERY
ANMP	ASSOCIAÇÃO NACIONAL DE MUNICÍPIOS PORTUGUESES
API	APPLICATION PROGRAMMING INTERFACE
AUC	AREA UNDER THE CURVE
CE	CRITÉRIO DE EXCLUSÃO
CI	CRITÉRIO DE INCLUSÃO
CISA	CYBERSECURITY AND INFRASTRUCTURE SECURITY AGENCY
CNCS	CENTRO NACIONAL DE CIBERSEGURANÇA
CPE	COMMON PLATFORM ENUMERATION
CPV	VOCABULÁRIO COMUM PARA OS CONTRATOS PÚBLICOS
CSV	COMMA-SEPARATED VALUES
CTI	CYBER THREAT INTELLIGENCE
CVE	COMMON VULNERABILITIES AND EXPOSURES
CVSS	COMMON VULNERABILITY SCORING SYSTEM
DBIR	DATA BREACH INVESTIGATIONS REPORT
DNS	DOMAIN NAME SYSTEM
DORA	DIGITAL OPERATIONAL RESILIENCE ACT
ENISA	AGÊNCIA DA UNIÃO EUROPEIA PARA A CIBERSEGURANÇA
EPSS	EXPLOIT PREDICTION SCORING SYSTEM
FICO	FAIR ISAAC CORPORATION
FIRST	FORUM OF INCIDENT RESPONSE AND SECURITY TEAMS
FN	FALSO NEGATIVO
FP	FALSO POSITIVO
GBDT	GRADIENT BOOSTED DECISION TREES
HSTS	HTTP STRICT TRANSPORT SECURITY
HTML	HYPERTEXT MARKUP LANGUAGE
HTTP	HYPERTEXT TRANSFER PROTOCOL
IA	INTELIGÊNCIA ARTIFICIAL
ICMP	INTERNET CONTROL MESSAGE PROTOCOL
IEC	INTERNATIONAL ELECTROTECHNICAL COMMISSION

IEEE	INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS
IMPIC	INSTITUTO DOS MERCADOS PÚBLICOS, DO IMOBILIÁRIO E DA CONSTRUÇÃO
IOC	INDICADOR DE COMPROMISSO
IOT	INTERNET DAS COISAS
IP	INTERNET PROTOCOL
IPAC	INSTITUTO PORTUGUÊS DE ACREDITAÇÃO
IRM	INSTITUTE OF RISK MANAGEMENT
ISO	INTERNATIONAL ORGANIZATION FOR STANDARDIZATION
ISO 27001	ISO/IEC 27001:2022 – INFORMATION SECURITY, CYBERSECURITY AND PRIVACY PROTECTION – INFORMATION SECURITY MANAGEMENT SYSTEMS – REQUIREMENTS
IT	TECNOLOGIAS DE INFORMAÇÃO
JSON	JAVASCRIPT OBJECT NOTATION
KEV	KNOWN EXPLOITED VULNERABILITIES
LIGHTGBM	LIGHT GRADIENT-BOOSTING MACHINE
LLM	LARGE LANGUAGE MODEL
MISP	MALWARE INFORMATION SHARING PLATFORM
NICSVAM	NON-INTRUSIVE CYBER SECURITY VULNERABILITY ASSESSMENT MODEL
NIF	NÚMERO DE IDENTIFICAÇÃO FISCAL
NINSRAPM	NON-INTRUSIVE NETWORK SECURITY RISK ASSESSMENT PREDICTION MODEL
NIPC	NÚMERO DE IDENTIFICAÇÃO DE PESSOA COLETIVA
NIS 2	NETWORK AND INFORMATION SECURITY DIRECTIVE 2
NIST	NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
NLP	PROCESSAMENTO DE LINGUAGEM NATURAL
NVD	NATIONAL VULNERABILITY DATABASE
ONU	ORGANIZAÇÃO DAS NAÇÕES UNIDAS
OSINT	OPEN-SOURCE INTELLIGENCE

OSIRIS-GOV	OPEN-SOURCE INTELLIGENCE FOR RESILIENT INFRASTRUCTURE AND SECURITY IN GOVERNMENT SERVICES
OT	TECNOLOGIAS OPERACIONAIS
OTX	OPEN THREAT EXCHANGE
PME	PEQUENAS E MÉDIAS EMPRESAS
PRISMA	PREFERRED REPORTING ITEMS FOR SYSTEMATIC REVIEWS AND META-ANALYSES
RF	REQUISITO FUNCIONAL
RNF	REQUISITO NÃO FUNCIONAL
ROC	RECEIVER OPERATING CHARACTERISTIC
RSS	REALLY SIMPLE SYNDICATION
SDK	SOFTWARE DEVELOPMENT KIT
SHOBEVODSDT	SHODAN AND BINARY EDGE BASED VULNERABLE OPEN DATA SOURCES DETECTION TOOL
SSL	SECURE SOCKETS LAYER
SURBL	SPAM URI REALTIME BLOCK LIST
TIC	TECNOLOGIAS DE INFORMAÇÃO E COMUNICAÇÃO
TLS	TRANSPORT LAYER SECURITY
UE	UNIÃO EUROPEIA
URI	UNIFORM RESOURCE IDENTIFIER
URL	UNIFORM RESOURCE LOCATOR
VCDB	VERIS COMMUNITY DATABASE
VERIS	VOCABULARY FOR EVENT RECORDING AND INCIDENT SHARING
VN	VERDADEIRO NEGATIVO
VP	VERDADEIRO POSITIVO
XGBOOST	EXTREME GRADIENT BOOSTING
XML	EXTENSIBLE MARKUP LANGUAGE

1. Introdução

Este capítulo contextualiza o trabalho a desenvolver na dissertação, expondo a motivação e o problema a resolver, que leva à definição dos objetivos para o projeto proposto. A par disto, apresentam-se as principais contribuições do mesmo e a estrutura deste documento.

1.1. Contexto, Motivação e Problema

Atualmente, a sociedade tem-se vindo a tornar cada vez mais digital [1]. Com o passar dos anos, o desenvolvimento da ciência e da tecnologia traz consigo uma inevitável transição digital, que suscita a conversão de processos anteriormente analógicos em novos processos digitais, tendo em vista o aproveitamento de todas as potencialidades das mais recentes inovações [2]. Por um lado, isto leva a uma crescente modernização da sociedade, inserindo a tecnologia nos vários momentos do quotidiano de forma quase ubíqua [1]. Por outro lado, esta nova realidade tem tendência a trazer associados novos riscos, com potencial para serem explorados por agentes maliciosos [3], [4].

De acordo com os mais recentes relatórios sobre ciberameaças, os números evidenciam precisamente esta tendência. A *Check Point Research* afirma que o número médio de ciberataques sofridos semanalmente por cada organização no segundo trimestre de 2025 (1984 ocorrências) aumentou 21% relativamente ao período homólogo do ano anterior (1636 ocorrências) e 58% em comparação com 2023 (1258 ocorrências) [5]. Analogamente, a Agência da União Europeia para a Cibersegurança (ENISA) mostra que o número de vulnerabilidades detetadas aumentou 27% no último ano, passando de 33524 (em 2024) para 42595 (em 2025) [3], [4]. Assim, estes indicadores confirmam que o presente traz progressivamente mais desafios para as organizações, que já deviam ter começado a colocar a cibersegurança como uma prioridade.

Para além desta necessidade de investir em cibersegurança devida aos riscos emergentes, habitualmente designada *risk-driven cybersecurity*, existe também a necessidade derivada da conformidade, ou *compliance-driven cybersecurity* [6], [7]. A este respeito, as principais normas, diretivas e regulamentos têm desempenhado um papel essencial, instando as organizações a adotarem múltiplas medidas no sentido de reforçar o seu nível de maturidade em cibersegurança [7], [8].

De facto, uma das principais normas internacionais de referência para a segurança da informação, publicada conjuntamente pela *International Organization for Standardization* (ISO) e pela *International Electrotechnical Commission* (IEC), a ISO/IEC 27001:2022 – *Information security, cybersecurity and privacy protection – Information security management systems – Requirements* (ISO 27001), tem vindo a ser cada vez mais implementada a nível nacional e internacional, conforme apontam os dados do Instituto Português de Acreditação (IPAC) e o *ISO Survey 2024*, respetivamente [9], [10], [11]. Mundialmente, o número de entidades certificadas mais do que duplicou entre 2023 (47291 organizações) e 2024 (96709 organizações), enquanto em Portugal este número também tem tendência a aumentar, com 162 entidades certificadas em 2022 e 252 em 2025, o que evidencia um crescente reconhecimento da importância atribuída à cibersegurança por parte das organizações [10], [11]. Com o intuito de implementar um sistema de gestão da segurança da informação, a ISO 27001 explicita 93 controlos de segurança no seu Anexo A, entre os quais se encontra, por exemplo, no A.5.19 *Information security in supplier relationships*, a implementação de processos e procedimentos concretos para gerir os riscos de segurança da informação associados ao uso de produtos ou serviços de fornecedores [9].

Ao nível do Direito da União Europeia (UE), a Diretiva (UE) 2022/2555 do Parlamento Europeu e do Conselho, de 14 de dezembro de 2022 – *Network and Information Security Directive* 2 (NIS 2) – transposta para a ordem jurídica nacional através do Decreto-Lei n.º 125/2025, de 4 de dezembro – estabelece medidas que visam reforçar a cibersegurança de forma transversal aos diferentes Estados-Membros [12], [13]. Em Portugal, encontram-se no âmbito deste novo regime jurídico as entidades consideradas essenciais, importantes ou públicas relevantes, sendo esta determinação efetuada conforme a sua dimensão e a criticidade do respetivo setor de atividade [13]. Neste sentido, as organizações abrangidas nestes três grupos devem cumprir as obrigações deste diploma, tais como (nos termos do disposto no artigo 21.º, número 2, alínea d) da diretiva original e, correspondentemente, no artigo 27.º, número 1, alínea c) da transposição nacional) a adoção de medidas que se foquem na "segurança da cadeia de abastecimento, incluindo aspetos de segurança respeitantes às relações entre cada entidade e os respetivos fornecedores ou prestadores de serviços diretos" [12], [13]. Em caso de incumprimento destas medidas, podem aplicar-se sanções que vão desde coimas até à suspensão de serviços ou à responsabilização individual dos titulares dos órgãos de gestão, pelo que as organizações têm sofrido uma forte pressão para assegurar a conformidade com estes requisitos [7], [12], [13].

Também na UE, o Regulamento (UE) 2022/2554 do Parlamento Europeu e do Conselho, de 14 de dezembro de 2022 – *Digital Operational Resilience Act* (DORA) –, obriga as entidades financeiras e os seus prestadores de serviços de Tecnologias de Informação e Comunicação (TIC) a medidas rigorosas, com o objetivo de aumentar a resiliência operacional digital do setor [14]. Este regulamento exige a identificação, avaliação, monitorização e gestão dos riscos da cadeia de fornecimento, dado que este vetor de ataque é cada vez mais explorado pelos cibercriminosos [14], [15].

Em síntese, todas estas três normas remetem para a implementação de medidas que visam prevenir ou mitigar o impacto dos ataques através da cadeia de fornecimento [9], [12], [13], [14]. Estes ataques consistem no comprometimento de um fornecedor ou prestador de serviços não como um fim em si mesmo, mas sim como um alvo intermédio, enquanto meio para atingir uma outra organização [15]. Para isto, contribui a diversidade de fornecedores que cada organização tem, por vezes deficientemente inventariados, bem como as reduzidas medidas de segurança por eles implementadas [15]. Entre julho de 2024 e junho de 2025, este tipo de ataques representou 10,6% do total de ciberincidentes registados pela ENISA, pelo que este é um vetor que deve ser considerado pelas organizações que procuram garantir a segurança dos seus ativos [4].

No setor da administração pública, maioritariamente abrangido pelo Decreto-Lei n.º 125/2025 de transposição da NIS 2 a nível nacional [13], os desafios de cibersegurança são particularmente prementes [3], [4]. Se ainda em 2024 este setor já era o alvo preferencial dos cibercriminosos, com 19% do total de incidentes observados pela ENISA, este valor tornou-se ainda mais substancial em 2025, período no qual a administração pública representou cerca de 38% do total de ciberataques reportados na UE [3], [4]. Portanto, tanto a nível central como local, estas organizações continuam a ser das principais impactadas pelo panorama atual de ciberameaças [3], [4].

Para se protegerem contra estas ciberameaças, as organizações do setor da administração pública podem recorrer à informação disponível publicamente relativa aos seus fornecedores e aos potenciais indicadores de risco associados à sua presença digital [16]. Efetivamente, o conceito de *Open-Source Intelligence* (OSINT) consiste em recolher e tratar os dados acessíveis e disponíveis ao público para os transformar em informação útil e inteligência acionável com uma determinada finalidade [16], [17]. Deste modo, a imensa quantidade de dados gerados pela crescente digitalização pode não só contribuir para a identificação das cadeias de fornecimento de cada entidade, como também aportar valor para a realização eficaz de uma avaliação do risco associado a cada fornecedor ou prestador de serviços [16].

No entanto, um dos principais obstáculos colocados às estratégias de OSINT deriva precisamente da enorme diversidade e heterogeneidade da informação passível de ser recolhida para ser analisada, o que pode requerer um extenso e moroso trabalho manual [17], [18]. Além disto, a automatização do processo de OSINT também acaba por ser um desafio, bem como a integração de diferentes fontes de informação e a extração de conhecimento a partir das mesmas, distinguindo o conteúdo relevante daquele que não o é [19]. Por isso, a inteligência artificial (IA) tem vindo a ser explorada como uma potencial forma de apoiar na resolução deste problema, embora não elimine, por si só, os desafios inerentes ao processo de OSINT [16]. Assim, importa considerar abordagens que, tendo em conta estas limitações, permitam estruturar e orientar o processo de OSINT de forma adequada ao contexto e ao objetivo em causa.

1.2. Objetivos

Esta dissertação pretende ir ao encontro deste problema e explorar as potencialidades de OSINT de maneira a suprir a necessidade que recai sobre as organizações no que concerne à correta identificação das suas cadeias de fornecimento e ao mapeamento do ciber-risco associado a cada organização. O foco deve incidir, essencialmente, no setor da administração pública, por ser um dos principais alvos das ciberameaças atuais. Portanto, o objetivo central passa por conceber e implementar um sistema de inteligência baseado em informação pública, capaz de identificar e monitorizar o risco digital associado a entidades da administração pública, tendo em conta os seus fornecedores e parceiros. Para esse efeito, deve ser mapeada automaticamente a cadeia de fornecimento observável das instituições públicas, englobando as relações diretas e indiretas de risco com terceiros.

Deste modo, podem ser detalhados os seguintes três objetivos específicos, tendo em vista o cumprimento do objetivo global do trabalho.

- Objetivo 1: Estudar o estado da arte relativamente a OSINT para recolher e tratar informação útil para avaliar o ciber-risco organizacional, incluindo o risco de terceiros e o seu potencial de propagação, tendo em vista a possibilidade de alcançar o principal objetivo proposto;
- Objetivo 2: Desenvolver uma ferramenta capaz de ir ao encontro do objetivo principal, isto é, mapear os fornecedores e prestadores de serviços das entidades do setor da administração pública, bem como o ciber-risco associado;
- Objetivo 3: Testar e validar a solução proposta.

Com o intuito de satisfazer estes objetivos, foi definida uma pergunta de investigação como ponto de partida para a pesquisa sistemática realizada: “Como é que a informação pública pode ser analisada para compreender o ciber-risco associado a uma entidade e às suas relações com terceiros?”. Esta questão suscitou o levantamento de diversas outras questões:

- Questão 1: Qual é o estado da arte relativamente à utilização de OSINT para compreender o ciber-risco de uma organização?
- Questão 2: Que informação é necessária para compreender o ciber-risco de uma organização?
- Questão 3: Que ferramentas existem para recolher e tratar esta informação?
- Questão 4: Quais são os principais desafios e limitações na utilização de OSINT para compreender o ciber-risco de uma organização?

1.3. Contribuições

O trabalho desenvolvido ao longo desta dissertação resultou nas seguintes três contribuições.

Em primeiro lugar, uma revisão de literatura para levantamento do estado da arte relativamente à avaliação do ciber-risco organizacional, tendo em consideração dois eixos: por um lado, a utilização de informação pública enquanto principal contributo para a aferição do ciber-risco e, por outro lado, o potencial impacto dos riscos de terceiros e da cadeia de fornecimento no ciber-risco avaliado. Esta análise permitiu comparar diferentes soluções existentes para ambos os problemas, identificando os seus principais desafios e limitações.

Em segundo lugar, uma solução – projetada e implementada – capaz de ir ao encontro das limitações identificadas anteriormente. Esta solução recolhe exclusivamente informação pública e considera os riscos de terceiros e da cadeia de fornecimento na avaliação do ciber-risco organizacional. Assim, com recurso a um classificador assente em aprendizagem automática supervisionada, são conjugados os dois domínios anteriores para uma avaliação mais completa do ciber-risco.

Em terceiro lugar, a aplicação da solução desenvolvida ao contexto da administração pública local portuguesa, enquanto validação da abordagem proposta num caso de estudo real.

1.4. Estrutura

Este documento está estruturado em seis capítulos.

O Capítulo 1 define o enquadramento no qual esta dissertação se insere, a motivação associada e o problema a resolver, bem como as principais contribuições alcançadas com o trabalho realizado.

O Capítulo 2 apresenta um levantamento do estado da arte em OSINT para avaliar o ciber-risco organizacional, considerando o risco de terceiros e o seu potencial de propagação, realizado através da análise de soluções e de trabalhos científicos para a identificação de falhas ou lacunas, através de uma metodologia de pesquisa sistemática para revisão da literatura.

O Capítulo 3 descreve a solução proposta para endereçar algumas das principais falhas, lacunas e limitações identificadas no levantamento do estado da arte, desde a definição de requisitos até à arquitetura delineada.

O Capítulo 4 contempla a implementação da solução proposta para o problema em questão, incluindo a apresentação de resultados perante as métricas consideradas relevantes para a avaliação da mesma.

O Capítulo 5 mostra a aplicação da solução desenvolvida ao caso de estudo da administração pública local portuguesa, exemplificando os resultados obtidos em determinadas câmaras municipais nacionais.

O Capítulo 6 conclui esta dissertação com as principais observações e os destaques essenciais do trabalho realizado, assim como com uma breve discussão de potenciais melhorias futuras a desenvolver.

2. Estado da Arte

Este capítulo contém uma revisão sistemática sobre o estado da arte nos temas relevantes para esta dissertação, de acordo com a metodologia *Preferred Reporting Items for Systematic reviews and Meta-Analyses* (PRISMA). Assim, expõem-se as principais ilações retiradas, em particular as falhas e lacunas atuais que o trabalho a desenvolver pretende endereçar.

2.1. Metodologia

O estudo realizado seguiu a metodologia PRISMA para pesquisa sistemática. Esta abordagem visa garantir que a revisão de literatura é efetuada de forma transparente, completa e rigorosa, com uma identificação clara das etapas seguidas e dos resultados alcançados, de maneira a assegurar que o estudo resultante tem tanto valor quanto possível [20].

A metodologia PRISMA é constituída por três fases: Identificação, Triagem e Inclusão. Em primeiro lugar, na fase de Identificação, efetua-se uma pesquisa em bases de dados científicas e noutras fontes de informação para identificar todos os registos (tais como artigos científicos, estudos e relatórios) potencialmente relevantes. Em segundo lugar, na fase de Triagem, procede-se a uma análise das publicações previamente identificadas para excluir resultados irrelevantes, tendo por base determinados critérios claramente definidos. Por último, na fase de Inclusão, restam todos os documentos que devem ser considerados na revisão sistemática de literatura para determinação do estado da arte [20], [21].

As subsecções seguintes detalham o processo seguido para a aplicação desta metodologia.

2.1.1. Fase de Identificação

Na fase de Identificação, o objetivo passou por identificar todos os registos potencialmente relacionados com o tema em estudo, através de uma pesquisa nas bases de dados científicas mais relevantes. Para esse efeito, definiram-se determinados domínios e palavras-chave, bem como as fontes a consultar.

De forma a garantir uma maior abrangência e eficácia da pesquisa, a mesma foi realizada em inglês, por ser este o idioma predominante na produção científica [22].

A principal pergunta de investigação anteriormente definida – "Como é que a informação pública pode ser analisada para compreender o ciber-risco associado a uma entidade e às suas relações com terceiros?" – abrange diversos domínios do conhecimento. Por isso, estes domínios foram previamente clarificados e estudados, tendo sido identificadas as palavras-chave consideradas mais relevantes para cada um deles.

Os domínios contemplados e as respetivas palavras-chave apresentam-se na Tabela 1.

Domínio	**Palavras-Chave**
Cibersegurança	"*cyber*", "*cybersecurity*"
Compreensão do Ciber-Risco	"*analysis*", "*assessment*", "*evaluation*", "*management*", "*scoring*"
Informação Pública	"OSINT", "*Open-Source Intelligence*", "*public information*", "*public data*", "*public sources*", "*publicly available*", "*non-intrusive*"
Risco	"*risk*"
Risco de Terceiros	"*third-party risk*", "*supply chain risk*", *"vendor risk*", *"risk cascading*", "*risk propagation*"

Tabela 1 – Domínios da Pergunta de Investigação

Conforme explicitado na Tabela 1, foram definidos cinco domínios associados à pergunta de investigação.

Em primeiro lugar, enquanto área do conhecimento mais abrangente na qual todos os restantes domínios se inserem, considerou-se o contexto global de cibersegurança, identificado pelo próprio termo ("*cybersecurity*") ou, frequentemente, apenas pelo seu prefixo ("*cyber*"), quando conjugado com outras palavras ou expressões relacionadas.

Em segundo lugar, o domínio relativo à compreensão do ciber-risco contemplou as palavras-chave associadas a este tema e vistas como mais relevantes, nomeadamente no que concerne à análise ("*analysis*") e avaliação ("*assessment*", "*evaluation*") do ciber-risco, bem como à atribuição de pontuações ("*scoring*") para facilitar a sua compreensão. Estes termos remetem para algumas das principais atividades relacionadas com a fase inicial do processo de gestão ("*management*") de ciber-riscos.

Em terceiro lugar, a ideia de extrair conhecimento a partir da informação disponível publicamente – enquanto elemento central da solução para o problema que se pretende endereçar – é mencionada na literatura científica através de diversas expressões com significados semelhantes, para além do conceito principal "OSINT" ou "*Open-Source Intelligence*". Por isso, consideraram-se como variantes destes termos as expressões "*public information*", "*public data*", "*public sources*", "*publicly available*" e "*non-intrusive*".

Para além disto, surgiu o domínio do risco, em referência ao ciber-risco, quando em coexistência com o primeiro. Este domínio encontra-se isolado dos restantes visto que um estudo preliminar da literatura mostrou que, frequentemente, o termo "*risk*" é combinado com outros termos de formas diferentes, nem sempre como "*cyber risk*" ou "*cybersecurity risk*", mesmo que em alusão ao ciber-risco. Portanto, optou-se pela separação dos domínios sobre cibersegurança e risco, mas assegurando a presença simultânea de ambos na expressão para consulta nas bases de dados, de maneira a remeter para o conceito de ciber-risco.

Finalmente, para englobar o ciber-risco associado às relações de uma organização com entidades externas, estabeleceu-se o último dos cinco domínios da Tabela 1. Este domínio contempla os riscos de terceiros, fornecedores e prestadores de serviços ("*third-party risk*", "*vendor risk*") e da cadeia de fornecimento ("*supply chain risk*"), bem como o estudo da sua propagação em cascata ("*risk cascading*", "*risk propagation*") entre entidades.

Deste modo, para agrupar todos estes domínios do conhecimento na expressão para consulta nas bases de dados de forma adequada, definiu-se a expressão da Figura 1.

```
("cyber" OR "cybersecurity")

AND

("analysis" OR "assessment" OR "evaluation" OR "management" OR "scoring")

AND

(

("risk" AND ("OSINT" OR "Open-Source Intelligence" OR "public information" OR "public data" OR "public sources" OR "publicly available" OR "non-intrusive"))

OR

("third-party risk" OR "supply chain risk" OR "vendor risk" OR "risk cascading" OR "risk propagation")

)
```

Figura 1 – Expressão para Consulta nas Bases de Dados

Tal como é visível na Figura 1, os domínios relacionados com cibersegurança e com a compreensão do ciber-risco são elementos nucleares da pesquisa e necessários em todos os registos procurados, daí encadeados de forma conjuntiva. No entanto, a relação entre a área de OSINT aplicado ao ciber-risco e o domínio do risco de terceiros é disjuntiva, visto que, embora se pretenda trabalhar exclusivamente com informação pública, a gestão de riscos de terceiros já é, tradicionalmente, realizada apenas com base nesta informação não intrusiva, pelo que forçar a inclusão de termos relacionados com OSINT seria limitativo em termos de resultados.

As bases de dados científicas selecionadas para consulta foram *Association for Computing Machinery* (ACM) *Digital Library*[1], *Institute of Electrical and Electronics Engineers* (IEEE) *Xplore*[2]*, Scopus*[3] e *Web of Science*[4], por serem das mais utilizadas nesta matéria.

[1] https://dl.acm.org/
[2] https://ieeexplore.ieee.org/
[3] https://www.scopus.com/
[4] https://www.webofscience.com/

Com o intuito de garantir maior relevância dos resultados, a expressão de pesquisa só incidiu sobre o título, o resumo e as palavras-chave de cada documento definidas pelos seus autores, adaptando-se a cada base de dados para cumprir este propósito.

A pesquisa – efetuada no dia 30 de novembro de 2025 nas quatro bases de dados científicas – retornou 701 resultados. Os resultados obtidos detalham-se na Tabela 2.

Base de Dados	**Número de Resultados**
ACM Digital Library	9
IEEE Xplore	91
Scopus	409
Web of Science	192
Total	**701**

Tabela 2 – Resultados da Pesquisa

Destes 701 resultados, 267 eram duplicados, isto é, foram retornados por mais do que uma base de dados, pelo que prosseguiram 434 para a fase de Triagem.

2.1.2. Fase de Triagem

Na fase de Triagem, analisaram-se os documentos resultantes da etapa anterior, de maneira a filtrar aqueles que poderiam, efetivamente, ser pertinentes para o estudo a realizar.

Cumprindo o propósito da metodologia PRISMA enquanto estratégia de pesquisa sistemática, rigorosa e transparente, a decisão de incluir ou excluir um determinado documento foi regida por critérios claros e bem definidos, delineados *a priori*, isto é, antes da leitura dos documentos.

Assim sendo, estabeleceram-se os critérios de inclusão (CI) da Tabela 3.

Critério	Descrição
CI1	O documento aborda a análise, avaliação, gestão ou pontuação do ciber-risco.
CI2	O documento recorre a informação pública ou a outros métodos não intrusivos para a recolha de dados.
CI3	O documento aborda o ciber-risco de uma entidade/organização, de terceiros, ou da cadeia de fornecimento.

Tabela 3 – Critérios de Inclusão

Analogamente, a Tabela 4 mostra os critérios de exclusão (CE).

Critério	Descrição
CE1	O documento tem uma abrangência setorial excessivamente específica, ao ponto de não ser generalizável (por exemplo, é apenas aplicável ao setor marítimo ou portuário, ao setor energético, etc.).
CE2	O documento cinge-se a avaliar perceções, opiniões ou tendências através de inquéritos, questionários, sondagens, ou abordagens semelhantes.
CE3	O documento contempla exclusivamente sistemas de controlo industrial, tecnologias operacionais (OT), sistemas ciber-físicos, *Internet* das Coisas (IoT), ou similares.
CE4	O documento é puramente conceptual, limitando-se a mapear conceitos de outros domínios ao contexto de cibersegurança.
CE5	O documento foca-se na implementação de controlos ou na adoção de medidas para mitigar ou tratar ciber-riscos.
CE6	O documento analisa a propagação do risco e o risco indireto apenas dentro de um sistema, em vez de entre organizações.
CE7	O documento recolhe informação exclusivamente a partir de fontes públicas já desatualizadas, sem manutenção ou atualmente inexistentes.

Tabela 4 – Critérios de Exclusão

Para ser incluído na revisão de literatura, cada documento teve de cumprir todos os critérios necessários para a sua inclusão e não encaixar em nenhum critério que obrigasse à sua exclusão. Por isso, segundo este princípio, os primeiros critérios – mais abrangentes – atuaram como filtro inicial de relevância, enquanto os segundos – mais específicos – permitiram excluir os documentos desalinhados do tema em estudo.

Com isto, procedeu-se, em primeiro lugar, à leitura do título e do resumo de cada um dos resultados previamente obtidos, com o intuito de avaliar a potencial relevância do texto completo. Assim, excluíram-se 386 registos, permanecendo 48.

Em segundo lugar, procuraram-se os documentos completos que passaram a etapa anterior. Neste caso, obtiveram-se 47 dos 48, por não ser possível aceder ao texto completo de um deles.

Finalmente, a triagem terminou com uma avaliação da elegibilidade de cada publicação, consistindo na leitura integral da mesma, tendo em vista uma apreciação mais precisa e fiável relativamente à sua utilidade para a revisão de literatura. Assim sendo, o resultado da aplicação de cada CE expõe-se na Tabela 5.

Critério	Número de Publicações Excluídas
CE1	5
CE2	3
CE3	2
CE4	2
CE5	1
CE6	1
CE7	1
Total	**15**

Tabela 5 – Número de Publicações Excluídas por Critério

Deste modo, restaram 32 publicações para a fase de Inclusão.

2.1.3. Fase de Inclusão

Nesta última fase, consolidaram-se os resultados obtidos.

Conforme exposto anteriormente, a metodologia PRISMA seguida a partir de bases de dados científicas levou à inclusão de 32 documentos.

No seguimento deste processo, após a leitura dos documentos incluídos, foi seguido um procedimento iterativo de consulta das referências dos mesmos, com o intuito de incluir outras publicações potencialmente relevantes para o levantamento do estado da arte. Este processo suscitou a inclusão de mais cinco registos.

Assim, resultou um total de 37 documentos incluídos na revisão de literatura.

O diagrama de fluxo da Figura 2 representa visualmente esta informação.

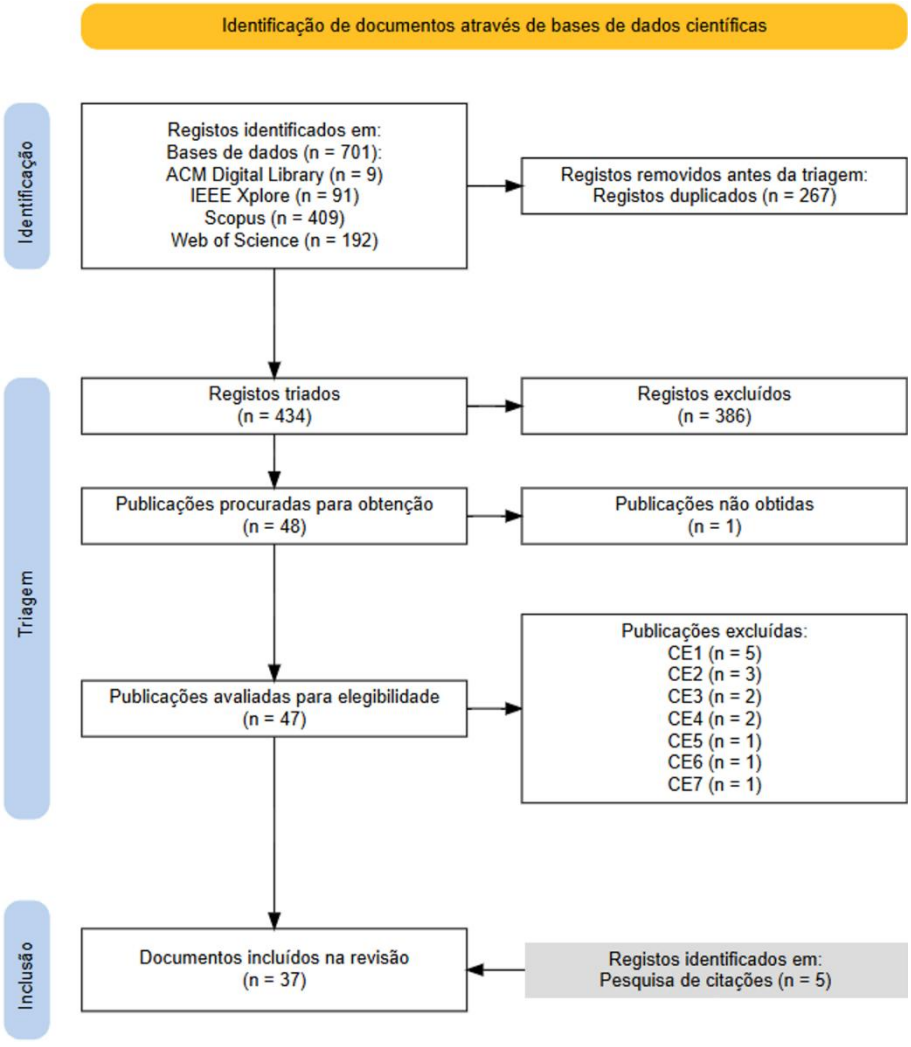

Figura 2 – Diagrama de Fluxo PRISMA

2.2. OSINT para Compreender o Ciber-Risco Organizacional

Para estudar o processo de OSINT orientado para a compreensão do ciber-risco organizacional, importa clarificar os conceitos de ciber-risco e processo de OSINT.

Para isso, o *National Institute of Standards and Technology* (NIST) dos Estados Unidos da América define o ciber-risco como "o risco de perda financeira, disrupção operacional ou dano a partir da falha das tecnologias digitais empregues para funções informacionais e/ou operacionais introduzido num sistema (…) através de mecanismos eletrónicos a partir de acesso não autorizado, uso, divulgação, disrupção, modificação ou destruição do sistema (…)" [23], ou, de forma mais simples, como "o risco de depender de ciber-recursos" [24], admitindo-se alguma sobreposição com os termos "risco de segurança", "risco de segurança da informação" e "risco de cibersegurança". Este termo mais lato abrange também os riscos devidos a ciberincidentes, a eventos de cibersegurança e a ataques no ciberespaço [23], [24]. De forma semelhante, o *Institute of Risk Management* (IRM) corrobora esta definição, mas simplifica-a ao "risco de perda financeira, disrupção ou dano à reputação de uma organização a partir de alguma falha dos seus sistemas de tecnologias de informação" [25].

Por sua vez, o processo de OSINT é considerado como a transformação de dados públicos em conhecimento, tendo em vista um determinado objetivo [16]. A maior parte dos autores estudados entende que este processo é constituído por cinco fases: (1) a identificação das fontes, (2) a recolha, (3) o tratamento e (4) a análise de dados e (5) a produção/apresentação de resultados [26].

Na primeira fase deste processo, são definidas as fontes de informação a consultar [26]. Estas fontes devem ser escolhidas conforme o objetivo pretendido, procurando que as informações disponibilizadas pelas mesmas sejam relevantes para o efeito [26], [27]. Posteriormente, na fase de recolha, consultam-se as fontes previamente definidas para se extrair informação, que é, depois, submetida a técnicas de filtragem, deduplicação e correlação, entre outras, na fase de tratamento [26], [28]. Após isto, na fase de análise, são processados os dados tratados tendo em vista a obtenção de conhecimento útil para o objetivo estabelecido [26], [28]. Estas três fases são, não raras vezes, agrupadas numa só [28]. Finalmente, na última fase do processo de OSINT, devem ser apresentados os resultados obtidos [26], [27].

Considerando esta estrutura do processo de OSINT, o estudo realizado detalha o estado da arte relativamente a cada uma das fases, tendo como objetivo a compreensão do ciber-risco organizacional.

2.2.1. Identificação das Fontes de Informação

O ponto de partida para o processo de OSINT passa pela identificação das fontes de informação pública a consultar, tendo em vista um determinado objetivo [26].

No contexto de cibersegurança, Sauerwein *et al.* [29] consideram que as fontes de informação relevantes para o processo de OSINT são aquelas que fornecem informação sobre vulnerabilidades, ameaças, ataques, riscos, ativos e contramedidas. Nesta medida, os autores definem uma vulnerabilidade como "uma fraqueza num ativo que pode ser explorada por uma ameaça" e uma ameaça como "a potencial causa de um incidente indesejado". Por sua vez, consideram um ataque como "qualquer tentativa não autorizada de aceder, alterar ou destruir um ativo", sendo um ativo "qualquer objeto ou característica que tenha valor para uma organização". Finalmente, veem uma contramedida como "qualquer controlo administrativo, de gestão, técnico ou legal que seja usado para contrariar um risco de segurança da informação", enquanto definem um risco como "as consequências de um potencial evento, como um ataque" [29].

Para além desta tipificação das fontes de informação, Sauerwein *et al.* [29] defendem que a capacidade de integração das mesmas é um aspeto essencial a considerar na delineação do processo de OSINT, por influenciar diretamente a possibilidade de automatização da recolha de dados. Por isso, os autores distinguem dois principais eixos: o formato e as interfaces disponíveis. Quanto ao formato, entendem que este pode ser estruturado/padronizado – como, a título de exemplo, *Comma-Separated Values* (CSV), *JavaScript Object Notation* (JSON), *HyperText Markup Language* (HTML) ou *eXtensible Markup Language* (XML) – ou não estruturado, ou seja, sem seguir um esquema de representação de dados comum e definido. Adicionalmente, atentam que a informação pode ser disponibilizada através de diversas interfaces. Em particular, são comuns os casos de *Application Programming Interfaces* (APIs) para este efeito, bem como *feeds Really Simple Syndication* (RSS), entre outros.

A par destes dois critérios, os autores destacam ainda três outros fatores que podem ser vistos como relevantes: a frequência de disponibilização da informação (regular ou irregular, ou seja, com periodicidade fixa ou apenas perante determinados eventos), a originalidade da mesma e a confiança/fiabilidade da fonte. Deste modo, Sauerwein *et al.* [29] entendem que também estes três parâmetros devem ser avaliados aquando da seleção das fontes de informação a consultar num processo de OSINT para compreensão do ciber-risco organizacional.

Assim, ainda que não referenciem diretamente esta taxonomia, as soluções que aplicam o processo de OSINT para compreender o ciber-risco organizacional acabam por ter em consideração estes fatores na escolha das fontes de informação. Em particular, o tipo de informação recolhida, bem como a interface e o formato, tendem a ser os aspetos aos quais é dada particular importância [30].

Quanto ao formato da informação disponibilizada, a tendência na literatura passa por preferir fontes que disponibilizem os seus dados de forma estruturada/padronizada, em detrimento de outras que não possuam esta característica. Esta opção deve-se ao facto de o formato fixo facilitar, naturalmente, a automatização do processo de recolha de informação [29], [30], [31]. Ainda assim, existem outras abordagens que optam pela recolha de dados não estruturados, por exemplo, para utilização em técnicas de Processamento de Linguagem Natural (NLP) [32].

No que concerne ao tipo de informação, a abordagem mais comum nos artigos estudados passa por recolher informação proveniente de fontes de diferentes categorias, cruzando, posteriormente, os resultados obtidos [30], [31], [33], [34]. Por exemplo, a estratégia mais usual das soluções que procuram avaliar o ciber-risco organizacional com base em informação pública consiste na recolha de informação sobre as vulnerabilidades que afetam os ativos corporativos expostos à *Internet*, tomando este como um indicador do ciber-risco [35], [36]. Além disto, também as fontes de informação sobre ataques e sobre ameaças são utilizadas noutras soluções [32], [37]. Excecionam-se, contudo, as fontes de informação sobre contramedidas e sobre riscos. Relativamente às primeiras, verifica-se que tendem a ser menos exploradas, uma vez que a informação sobre as contramedidas adotadas por uma organização é, por norma, interna, o que limita a sua disponibilização pública de forma credível e verificável [35]. Quanto às segundas, a informação sobre os riscos costuma ser obtida através do cruzamento e da combinação dos resultados provenientes de outras fontes, conforme já referido [30], [34].

Deste modo, são particularmente relevantes as fontes de informação relativas a vulnerabilidades, ativos, ataques e ameaças. Para cada categoria, deve ser considerada a interface e o formato da informação disponibilizada, enquanto informações úteis para a delineação do processo de OSINT.

Com este intuito, a Tabela 6 elenca as fontes públicas de segurança da informação sobre vulnerabilidades referenciadas pelos artigos analisados e atualmente disponíveis.

Fonte	Interface	Formato	Referência
Common Vulnerabilities and Exposures (CVE) *Details*[5]	API, RSS	JSON, XML	*[29]*
CVE Program[6]	API	JSON	*[29]*
CXSECURITY.COM[7]	RSS	XML	*[29]*
Cybersecurity and Infrastructure Security Agency (CISA) *Known Exploited Vulnerabilities* (KEV)[8]	*Web*	CSV, JSON	*[30], [38]*
Forum of Incident Response and Security Teams (FIRST) *Exploit Prediction Scoring System* (EPSS)[9]	API	JSON	*[30], [38]*
National Vulnerability Database (NVD)[10]	API	JSON	*[29], [30], [32], [35], [36], [38]*
Security Database[11]	API	JSON	*[29]*
VulDB[12]	API, RSS	JSON, XML	*[29]*
Vulners[13]	API	JSON	*[29]*

Tabela 6 – Fontes de Informação sobre Vulnerabilidades

[5] https://www.cvedetails.com/
[6] https://www.cve.org/
[7] https://cxsecurity.com/
[8] https://www.cisa.gov/known-exploited-vulnerabilities-catalog
[9] https://www.first.org/epss/
[10] https://nvd.nist.gov/
[11] https://www.security-database.com/
[12] https://vuldb.com/
[13] https://vulners.com/

A Tabela 6 destaca nove fontes de informação sobre vulnerabilidades, consideradas pelos diferentes autores para recolher informação capaz de contribuir para a aferição do ciber-risco que afeta as organizações.

O *CVE Details* é um repositório parcialmente aberto, propriedade da *SecurityScorecard*[14], que cruza informação de múltiplas outras fontes para proporcionar uma visão agregada e contextualizada das vulnerabilidades. Esta plataforma é levantada como uma opção para recolher informação sobre vulnerabilidades por Sauerwein *et al.* [29], destacando-se o facto de disponibilizar os dados tanto em formato JSON, através de uma API, quanto em XML, por via de um *feed* RSS. No entanto, nenhum dos autores analisados acaba por recorrer a esta fonte de informação, uma vez que, em comparação com outras alternativas, é limitada na informação que disponibiliza de forma gratuita e aberta ao público.

O *CVE Program* é o programa internacional gerido pela MITRE[15] que atribui um identificador único (no formato "CVE-YYYY-NNNN", onde YYYY representa o ano de atribuição do identificador à vulnerabilidade e NNNN é um número sequencial) a cada vulnerabilidade conhecida. Esta fonte funciona como um repositório centralizado de vulnerabilidades ao qual é atribuída a máxima credibilidade, por ser patrocinado pela CISA. Assim sendo, Sauerwein *et al.* [29] sugerem a consulta desta fonte de informação para confirmar e corroborar as informações obtidas noutras fontes sobre vulnerabilidades, de maneira a garantir a veracidade dos dados e a minimizar ou evitar eventuais falsos positivos. Contudo, na prática, tende a ser dada preferência a outras fontes de informação sobre vulnerabilidades, por disponibilizarem informação com a mesma credibilidade e facilidade de consulta, mas com mais contexto e com métricas capazes de oferecer um melhor contributo para a aferição do ciber-risco, como é o caso da NVD.

O *CXSECURITY.COM* é uma plataforma pública que recolhe e divulga informações sobre vulnerabilidades e *exploits*, maioritariamente reportadas pela comunidade de cibersegurança internacional. No entanto, embora seja suscitada como uma hipótese por Sauerwein *et al.* [29], não tem implementações práticas na literatura estudada, provavelmente por não ser uma fonte tão credível e fiável como as restantes, tendo em conta que muitas das informações disponibilizadas provêm de publicações de utilizadores anónimos.

[14] https://securityscorecard.com/
[15] https://www.mitre.org/

A CISA KEV é uma base de dados que lista as vulnerabilidades conhecidas que estão ativamente a ser exploradas por atacantes. Tendo em conta a elevada frequência com que surgem novas vulnerabilidades, esta base de dados assume um papel preponderante na recolha de informação pertinente para avaliar o nível de risco de uma organização, ao permitir distinguir as vulnerabilidades dificilmente exploráveis daquelas que estão, comprovadamente, a ser aproveitadas para finalidades maliciosas. Assim, a CISA KEV aporta um importante contributo por possibilitar a priorização das vulnerabilidades detetadas com base nas evidências conhecidas de exploração das mesmas. Esta informação é utilizada de dois modos distintos na literatura recolhida. Por um lado, Babenko *et al.* [30] recorrem à base de dados CISA KEV para validar, *a posteriori*, as previsões resultantes do modelo de aprendizagem automática desenvolvido para classificação do ciber-risco organizacional, funcionando, portanto, como um instrumento de avaliação e comparação com a realidade do ciber-risco calculado perante cada vulnerabilidade. Por outro lado, Harry *et al.* [38] comparam as vulnerabilidades presentes na base de dados CISA KEV com as que são identificadas nos ativos expostos à *Internet* das entidades analisadas, considerando que, em caso de correspondência, tal constitui um fator agravador do ciber-risco da organização.

O FIRST EPSS é um sistema de pontuação que procura estimar a probabilidade de exploração de uma vulnerabilidade, atribuindo a cada vulnerabilidade um valor entre zero (probabilidade mínima de exploração) e um (probabilidade máxima de exploração). Assim, de forma semelhante à fonte de informação anterior, esta tem o intuito de facilitar a priorização de diferentes vulnerabilidades, mas consoante o seu potencial de exploração. Deste modo, a diferença entre a base de dados CISA KEV e o FIRST EPSS prende-se com o facto de a primeira ser uma fonte de informação que lista as vulnerabilidades que, comprovadamente, se encontram a ser exploradas por agentes maliciosos, enquanto o segundo fornece apenas uma estimativa da probabilidade de exploração de cada vulnerabilidade. Na literatura, Babenko *et al.* [30] utilizam a informação do FIRST EPSS de forma similar à proveniente da CISA KEV, ou seja, como forma de validação dos resultados do seu modelo de aprendizagem automática. Diferentemente, Harry *et al.* [38] recolhem o valor do FIRST EPSS para cada vulnerabilidade detetada – enquanto estimador da probabilidade de exploração – e multiplicam-no pela respetiva pontuação na escala *Common Vulnerability Scoring System* (CVSS) – enquanto estimador da severidade, ou seja, do impacto em caso de exploração – para obter um indicador do ciber-risco que contempla, simultaneamente, a probabilidade de exploração da vulnerabilidade e o impacto da mesma.

A NVD é a base de dados pública mantida pelo NIST que enriquece cada CVE com informação técnica contextual e métricas sobre o seu potencial impacto, como o respetivo CVSS. Em comparação com o *CVE Program*, esta fonte é considerada igualmente credível e com a mesma facilidade de consulta via API, mas tende a ser preferida devido ao maior contexto que atribui a cada vulnerabilidade. Enquanto fonte de informação primária de referência para as vulnerabilidades, a NVD é proposta por Sauerwein *et al.* [29] e utilizada pela vasta maioria dos autores estudados para recolha de informação sobre vulnerabilidades. Na literatura, Babenko *et al.* [30] extraem desta fonte de informação as versões de *software* afetadas por cada vulnerabilidade, para cruzarem com as versões de *software* identificadas em cada organização. Noutra perspetiva, Silvestri *et al.* [32] utilizam o catálogo de *Common Platform Enumerations* (CPEs) – enquanto esquema padronizado para identificação de aplicações, *hardware*, *software* e sistemas de IT – disponibilizado pela mesma fonte para mapear cada ativo organizacional numa determinada categoria consoante as suas funcionalidades, de modo a classificar o respetivo nível de criticidade. Finalmente, Ribeiro *et al.* [35], Genge *et al.* [36] e Harry *et al.* [38] recorrem a estes mesmos dados para efetuar o cruzamento entre as vulnerabilidades detetadas e as versões dos sistemas das organizações. Todos estes autores extraem o valor da severidade de cada vulnerabilidade detetada nos ativos corporativos, mas Silvestri *et al.* [32] diferenciam-se dos restantes por considerarem também o relatório textual associado, que submetem a técnicas de NLP para melhor contextualização do impacto de exploração da vulnerabilidade.

Finalmente, a *Security Database*, o *VulDB* e o *Vulners* são plataformas comerciais agregadoras de vulnerabilidades, maioritariamente focadas na integração com outros programas de gestão de vulnerabilidades. Embora sejam referenciadas por Sauerwein *et al.* [29], as limitações existentes nas versões gratuitas destas plataformas levam a que não sejam adotadas pelos autores estudados.

Conforme apontado anteriormente, a extração de informação pertinente para compreender o ciber-risco de uma entidade a partir dos dados de vulnerabilidades requer conhecimento dos ativos organizacionais, na medida em que a existência de uma vulnerabilidade só tem um contributo agravador do ciber-risco de uma organização se afetar algum dos seus ativos. Caso contrário, a relevância das vulnerabilidades enquanto indicadores isolados é vista como residual. Portanto, as soluções que integram informação sobre vulnerabilidades tendem a complementar a sua análise com fontes que permitam identificar os ativos da organização, possibilitando o cruzamento entre ambos os dados [30], [35], [36].

Para isso, existem fontes que atuam como motores de pesquisa de dispositivos visíveis, endereços *Internet Protocol* (IP), serviços expostos, etc., pelo que permitem identificar os ativos de uma organização. A Tabela 7 apresenta as principais fontes para este efeito, referidas pelas soluções apresentadas nos artigos estudados.

Fonte	Interface	Formato	Referência
Censys[16]	API	JSON	*[30], [31], [35], [39]*
FOFA[17]	API	JSON	*[34]*
Shodan[18]	API	JSON	*[30], [31], [34], [35], [36], [39]*
WHOIS[19]	*Web*	N/A	*[30], [31]*
ZoomEye[20]	API	JSON	*[31]*

Tabela 7 – Fontes de Informação sobre Ativos

As fontes elencadas na Tabela 7 disponibilizam informação sobre os ativos de cada organização, sendo o *WHOIS* a mais distinta das demais.

O *Censys* é uma das plataformas internacionalmente mais utilizadas enquanto motor de pesquisa de ativos expostos à *Internet*. A par desta informação, também disponibiliza dados detalhados sobre certificados *Secure Sockets Layer* (SSL)/ *Transport Layer Security* (TLS), sendo, por isso, usada por Babenko *et al.* [30] como referência comparativa e de validação do sistema desenvolvido para recolha de informação sobre estes certificados. Para além de aproveitarem esta funcionalidade, Ribeiro *et al.* [35] e Lee e Shon [39] recorrem ao *Censys* para descobrirem subdomínios, endereços IP, portas expostas e serviços em execução nas organizações, tendo em vista o mapeamento da sua superfície de ataque. Por fim, Daskevics e Nikiforova [31], embora não utilizem diretamente esta fonte, referem-na como possível implementação futura da solução por eles desenvolvida para aferição do ciber-risco organizacional.

[16] https://platform.censys.io/
[17] https://en.fofa.info/
[18] https://www.shodan.io/
[19] https://who.is/
[20] https://www.zoomeye.ai/

O *FOFA* é uma plataforma semelhante à anterior, mas com principal prevalência no continente asiático. Por isso, Yang *et al.* [34] tiram proveito desta fonte de informação para recolher dados sobre os ativos das organizações, de maneira a mapear a respetiva superfície de ataque e, com isso, ser capaz de compreender o ciber-risco a que estão sujeitas, por via da exposição dos seus ativos à *Internet*.

O *Shodan* é, a par do *Censys*, a fonte de referência para obter informações sobre ativos expostos à *Internet*. Por isso mesmo, é consultada por diversos autores. Babenko *et al.* [30] utilizam-na, por um lado, para automatizar a descoberta de ativos expostos à *Internet* através de pedidos via API e, por outro lado, para comparar os relatórios de cobertura da superfície de ataque disponibilizados pelo *Shodan* com os gerados pela sua ferramenta. Por sua vez, Daskevics e Nikiforova [31] filtram a pesquisa que efetuam no *Shodan* para obterem apenas endereços IP correspondentes a bases de dados expostas à *Internet*, com as quais se tenta, posteriormente, estabelecer uma ligação para aferir o ciber-risco associado. Num âmbito mais abrangente, Yang *et al.* [34] procuram identificar portas e serviços expostos pelas organizações, bem como outras informações relativas à segurança das suas aplicações *web*. A par disto, Ribeiro *et al.* [35] e Genge *et al.* [36] pesquisam pelos ativos organizacionais expostos à *Internet* através dos seus domínios e endereços IP, de modo a obter as versões de *software* dos sistemas em execução nos mesmos, tendo em vista o cruzamento desses dados com as informações sobre vulnerabilidades anteriormente mencionadas. Finalmente, Lee e Shon [39] tratam o *Shodan* da mesma forma que o *Censys*, ou seja, para mapear a superfície de ataque de cada organização.

De forma diferente das fontes anteriores, o *WHOIS* permite recolher informações associadas a domínios e subdomínios, nomeadamente registos *Domain Name System* (DNS) e certificados SSL/TLS, em vez de dados sobre ativos expostos à *Internet*. Embora estas informações não sejam, nos artigos referenciados, tão recorrentes quanto as anteriores para a compreensão do ciber-risco organizacional, nalguns casos também acabam por ser contempladas em metodologias de avaliação do ciber-risco. Em particular, Babenko *et al.* [30] e Daskevics e Nikiforova [31] recorrem a estes dados numa análise preliminar da organização em estudo, de maneira a identificar potenciais alvos para serem submetidos às outras fontes de informação deste tipo. Como esta fonte de informação não disponibiliza os dados através de uma API, nem de forma padronizada/estruturada, os autores acabam por recorrer a outras fontes similares que disponibilizam a mesma informação, mas em formato JSON ou XML, via API.

Por último, o *ZoomEye* é, de forma análoga ao *FOFA*, outra das fontes de informação sobre ativos com maior utilização na Ásia. Ainda assim, acaba por não ser, na prática, utilizada por nenhum dos autores recolhidos, sendo apenas referida por Daskevics e Nikiforova [31] como uma potencial fonte de informação a consultar em trabalhos futuros.

Além da informação que pode ser obtida a partir da consulta destas fontes de informação sobre vulnerabilidades e ativos, também os ataques ou incidentes concretizados podem resultar em informação que pode vir a ser disponibilizada publicamente. Por um lado, as fugas de dados são analisadas por alguns autores para aferir o ciber-risco organizacional [40], [41] e, por outro lado, também os padrões estatísticos de incidentes são considerados por outros [37], [42].

Nesta medida, os artigos lidos recorrem às fontes de informação sobre ataques da Tabela 8.

Fonte	Interface	Formato	Referência
Have I Been Pwned[21]	API	JSON	*[40], [41]*
Intelligence X[22]	API	CSV, JSON	*[40]*
Vocabulary for Event Recording and Incident Sharing (VERIS) *Community Database* (VCDB)[23]	*Web*	JSON	*[37], [42]*

Tabela 8 – Fontes de Informação sobre Ataques

Embora todas as três fontes da Tabela 8 forneçam informação sobre ataques passados, têm propósitos e âmbitos distintos.

[21] https://haveibeenpwned.com/
[22] https://intelx.io/
[23] https://verisframework.org/vcdb.html

O *Have I Been Pwned* tem como principal foco a disponibilização de endereços de *e-mail* e de palavras-passe comprometidas e divulgadas em fugas de dados. Por isso, Mukhopadhyay e Luther [40] e Hayes e Cappa [41] propõem a utilização desta fonte de informação para verificar a exposição associada a endereços de *e-mail* corporativos, considerando este um indicador do ciber-risco organizacional.

Similarmente, o *Intelligence X* é um motor de pesquisa que também incide sobre os dados provenientes de fugas de dados, mas acaba por ser mais abrangente, ao indexar múltiplos outros documentos originários de máquinas infetadas e comprometidas, para além de endereços de *e-mail* e palavras-passe. Tal como anteriormente, Mukhopadhyay e Luther [40] levantam a possibilidade de consultar esta fonte de informação para compreender o ciber-risco das organizações de forma não intrusiva, através da sua exposição digital.

Diferentemente, a VCDB é um repositório de incidentes de segurança reportados publicamente e estruturados no formato utilizado pela VERIS para a elaboração do seu relatório anual *Data Breach Investigations Report* (DBIR), que contém indicadores demográficos sobre as organizações vítimas destes ataques. Liu *et al.* [37] cruzam esta informação com outros indicadores para desenvolver um modelo de aprendizagem automática com o objetivo de prever futuros ciberincidentes, enquanto Sarabi *et al.* [42] extraem as características das organizações desta base de dados na tentativa de identificar tendências nos respetivos dados demográficos, de modo a criar um modelo de aprendizagem automática com o mesmo objetivo.

Finalmente, restam as fontes de informação sobre ciberameaças. Por um lado, através da inteligência obtida em processos de *Cyber Threat Intelligence* (CTI), é relevante ter conhecimento das principais ameaças que incidem não só sobre uma organização em particular, mas também sobre o país ou o setor em que se insere [29], [43]. Por outro lado, estas fontes também podem ser úteis para, ao cruzar a informação com os dados previamente mencionados, identificar potenciais ameaças desconhecidas que estejam associadas ao comprometimento de ativos da organização. Por exemplo, a presença de um endereço IP de uma organização numa lista de endereços IP relacionados com atividade maliciosa, ou a associação de domínios corporativos a campanhas de distribuição de *phishing* ou *malware*, são fatores que alguns autores entendem ter impacto no ciber-risco de uma organização [29], [34], [37].

A Tabela 9 mostra as diferentes fontes utilizadas na literatura para obter este tipo de informações.

Fonte	Interface	Formato	Referência
AbuseIPDB[24]	API	JSON	*[33], [34]*
AlienVault Open Threat Exchange (OTX)[25]	N/A	N/A	*[29]*
DShield[26]	API	CSV, JSON, XML	*[29], [37]*
Malware Information Sharing Platform (MISP)[27]	N/A	N/A	*[29], [43]*
PhishTank[28]	API	CSV, JSON, XML	*[29], [37]*
SpamCop[29]	API	JSON	*[37]*
Spamhaus[30]	API	JSON	*[37]*
Spam Uniform Resource Identifier (URI) *Realtime Block List* (SURBL)[31]	API	CSV, JSON	*[37]*
VirusTotal[32]	API	JSON	*[29], [31]*

Tabela 9 – Fontes de Informação sobre Ameaças

Apesar de, genericamente, as fontes listadas na Tabela 9 fornecerem informação sobre ameaças para compreender o risco de uma organização, os seus propósitos e modos de operação são distintos, assim como a forma e as razões pelas quais são utilizadas nos artigos incluídos.

[24] https://www.abuseipdb.com/
[25] https://otx.alienvault.com/
[26] https://www.dshield.org/
[27] https://www.misp-project.org/
[28] https://www.phishtank.com/
[29] https://www.spamcop.net/
[30] https://www.spamhaus.org/
[31] https://www.surbl.org/
[32] https://www.virustotal.com/

O *AbuseIPDB* é orientado para avaliar endereços IP que estejam a veicular atividades potencialmente maliciosas. Assim sendo, Yang *et al.* [33], [34] recorrem a esta fonte para verificar se os endereços IP das organizações se encontram associados a comportamentos maliciosos, para que, em caso afirmativo, seja refletido esse facto em indicadores não intrusivos de avaliação do ciber-risco organizacional.

O *AlienVault OTX* vai além do sentido tradicional de fonte, na medida em que permite não só a recolha, mas também a partilha de informação entre diferentes entidades, em particular sobre Indicadores de Compromisso (IoC) e outras informações relacionadas com inteligência sobre ciberameaças. Nesta medida, é sugerido por Sauerwein *et al.* [29] como uma possível fonte pública de segurança da informação, mas não encontra aplicação na literatura analisada.

Tal como o *AbuseIPDB*, o *DShield* reporta endereços IP relacionados com atividade maliciosa. Por isso, também é levantada por Sauerwein *et al.* [29] enquanto fonte pública de segurança da informação e é aproveitada por Liu *et al.* [37] para contabilizar os endereços IP de cada organização presentes nesta base de dados, assim como a sua variação ao longo do tempo, sendo estes valores tomados como atributos do modelo de aprendizagem automática desenvolvido para a avaliação do ciber-risco.

De forma idêntica ao *AlienVault OTX*, o MISP é uma plataforma de partilha de informação sobre ciberameaças, tradicionalmente utilizada no contexto de CTI. Sucintamente, também permite recolher (e submeter) inteligência sobre ciberameaças, que pode contribuir para determinar o ciber-risco associado a uma organização. No entanto, apesar de ser elencada como uma possibilidade por Sauerwein *et al.* [29], a única implementação na literatura recolhida encontra-se na solução de Agostinho *et al.* [43], que não visa compreender o ciber-risco organizacional de forma proativa e não intrusiva, mas apenas contribuir para a contextualização reativa de ameaças concretas já identificadas.

O *PhishTank* foca-se em identificar domínios relacionados com a divulgação de campanhas de *phishing*. Deste modo, para além de ser suscitada como uma possibilidade por Sauerwein *et al.* [29], é consultada por Liu *et al.* [37] da mesma forma que o *DShield*, isto é, para medir o número de ocorrências de domínios das organizações nesta base de dados e a sua variabilidade temporal, como métricas a submeter ao mesmo modelo de aprendizagem automática para avaliar o ciber-risco.

O *SpamCop*, o *Spamhaus* e o SURBL são mais vocacionados para a deteção de endereços IP, URIs ou domínios associados à propagação de *spam*. Igualmente, todas estas fontes são consultadas – com o mesmo intuito das anteriores – por Liu *et al.* [37], sendo o número de presenças dos ativos das organizações nas mesmas e a respetiva tendência de crescimento ou decrescimento temporal tratados como atributos do modelo desenvolvido pelos autores.

Por último, o *VirusTotal* é uma fonte de informação mais transversal, que agrega resultados de múltiplos antivírus e de outras fontes para analisar ficheiros, *Uniform Resource Locators* (URLs) e endereços IP, com uma inteligência mais abrangente. Nesta medida, é listada por Sauerwein *et al.* [29] e referida por Daskevics e Nikiforova [31] como uma possível solução para consultar informação sobre as organizações a analisar.

Em resumo, existe uma vasta panóplia de fontes que podem atuar como ponto de partida para a recolha de informação no processo de OSINT, tendo como objetivo a compreensão do ciber-risco que impende sobre uma organização.

A literatura analisada demonstra que as soluções existentes tendem a privilegiar fontes que disponibilizam os dados de forma estruturada e acessíveis via API, uma vez que estas facilitam a automatização da recolha dos mesmos. De forma geral, a maior parte das abordagens cruza informação sobre ativos expostos à *Internet* com informação sobre vulnerabilidades, procurando identificar se os sistemas associados a uma organização se encontram afetados por vulnerabilidades conhecidas. Em paralelo, alguns trabalhos recorrem também a fontes de informação sobre ameaças, analisando se os ativos organizacionais apresentam sinais de comprometimento, detetados através de indícios de atividade potencialmente maliciosa.

Contudo, a literatura evidencia uma forte concentração na análise da superfície de ataque técnica das organizações. As fontes de informação sobre ataques já concretizados, nomeadamente aquelas que revelam a exposição de dados das organizações, têm uma menor presença nos trabalhos analisados, apesar de as evidências recentes apontarem o comprometimento de credenciais encontradas em fugas de dados como um dos vetores de ataque mais frequentes [44]. Esta constatação sugere a relevância de se considerar este tipo de informação enquanto indicador de exposição digital, bem como de compreender de que forma é que a informação pode ser recolhida, tratada e analisada no contexto de processos de OSINT orientados para a compreensão do ciber-risco organizacional.

2.2.2. Recolha, Tratamento e Análise de Informação

No processo de OSINT, as fases de recolha, tratamento e análise de informação são, frequentemente, agrupadas numa só, visto que muitas ferramentas acabam por as realizar de forma integrada, recolhendo os dados, tratando-os e analisando-os [19].

Estas ferramentas podem ser de propósito geral – com utilidade para vários objetivos, mas aplicáveis à avaliação do ciber-risco organizacional – ou desenvolvidas especificamente para esta finalidade.

A Tabela 10 sintetiza as ferramentas de propósito geral utilizadas nos artigos analisados, com destaque para as fontes de informação que cada uma delas integra, de entre as referidas anteriormente.

Ferramenta	Fontes	Referência
amass[33]	*AlienVault OTX*, *FOFA*, *Intelligence X*, *Shodan*, *VirusTotal*, *WHOIS*, *ZoomEye*	*[30]*
Maltego[34]	*AbuseIPDB*, *AlienVault OTX*, *Censys*, *Have I Been Pwned*, MISP, NVD, *Shodan*, *Spamhaus*, *VirusTotal*, *Vulners*, *WHOIS*	*[19], [41]*
SpiderFoot[35]	*AbuseIPDB*, *AlienVault OTX, Censys, Have I Been Pwned, Intelligence X*, *PhishTank*, *Shodan*, *SpamCop*, *Spamhaus*, SURBL, *VirusTotal*, *WHOIS*	*[19], [30]*
Sudomy[36]	*AlienVault OTX*, *Censys*, *Shodan*, *VirusTotal*	*[31]*
theHarvester[37]	*AlienVault OTX*, *Censys*, *FOFA*, *Have I Been Pwned*, *Intelligence X*, *Shodan*, *VirusTotal*, *WHOIS*, *ZoomEye*	*[39]*

Tabela 10 – Ferramentas de Propósito Geral

[33] https://github.com/owasp-amass/amass
[34] https://www.maltego.com/
[35] https://www.spiderfoot.net/
[36] https://github.com/screetsec/Sudomy
[37] https://github.com/laramies/theHarvester

As cinco ferramentas de propósito geral da Tabela 10 são, maioritariamente, utilizadas para efeitos de mapeamento da superfície de ataque, mas alguns autores conseguem aproveitá-las para efeitos de compreensão do ciber-risco organizacional.

O *amass* é uma ferramenta *open-source* que visa mapear a superfície de ataque e descobrir ativos das organizações, recolhendo os dados a partir de fontes públicas das elencadas anteriormente. Nesta medida, constitui um possível ponto de partida para compreender o ciber-risco de uma organização, com base na identificação da superfície exposta e das suas potenciais vulnerabilidades ou de eventuais comprometimentos. Na literatura, é utilizado por Babenko *et al.* [30] para fins de validação dos resultados e comparação do desempenho da solução desenvolvida pelos autores, destacando-se a capacidade para processar cerca de 50.000 domínios por hora, através de diversas interações com as APIs das fontes de informação consultadas.

Noutra perspetiva, o *Maltego* é uma das ferramentas mais usuais para realizar o mapeamento de relações entre entidades e possibilitar a sua visualização gráfica, integrando diversas fontes [19]. Por isso, esta ferramenta é essencialmente vocacionada para a realização de investigações, ao permitir o desenho de grafos que evidenciem as interligações entre diferentes elementos. No entanto, o *Maltego* encontra, no artigo de Hayes e Cappa [41], uma utilização para aferir o ciber-risco no que concerne à propensão a ataques de engenharia social, avaliada através da identificação da exposição digital dos colaboradores de uma organização, que é determinada com o auxílio de grafos traçados pelo *Maltego* a partir dos respetivos endereços de *e-mail*.

Por sua vez, o *SpiderFoot* é uma das ferramentas *open-source* de propósito geral mais amplamente utilizadas no contexto de OSINT aplicado à cibersegurança [19]. Esta ferramenta automatiza a consulta, correlação e análise de dados provenientes de mais de 200 fontes públicas, entre as quais se incluem, por exemplo, os já mencionados *Censys* e *Shodan*, bem como *Have I Been Pwned*, *VirusTotal* e *WHOIS*, entre outros. Com isto, o *SpiderFoot* é capaz de monitorizar a exposição da infraestrutura tecnológica de uma organização de forma integrada numa só plataforma. Tendo em vista o aproveitamento da versatilidade do *SpiderFoot*, Babenko *et al.* [30] utilizam os resultados obtidos por esta ferramenta como termo de comparação para a sua solução de avaliação do ciber-risco organizacional, principalmente em termos de desempenho. Nesta medida, os autores destacam que o *SpiderFoot* assenta numa arquitetura modular e que a recolha, a análise e o tratamento de dados por parte desta ferramenta são orientados por regras, em vez de seguirem uma abordagem baseada em IA, tendo uma capacidade de processamento de cerca de 10.000 organizações por hora.

De forma similar ao *amass*, também o *Sudomy* é uma ferramenta *open-source* de reconhecimento e mapeamento da superfície de ataque, mas particularmente orientada para a enumeração de domínios e subdomínios. Esta ferramenta é considerada e avaliada no artigo de Daskevics e Nikiforova [31] para efeitos de compreensão do ciber-risco organizacional da forma pretendida pelos autores, mas acaba por ser vista como insuficiente dadas as limitações do seu âmbito, que se cinge, essencialmente, à enumeração de domínios. Por isso, estas mesmas limitações suscitam o desenvolvimento de uma outra ferramenta pelos próprios autores para os efeitos desejados, explicada posteriormente.

Finalmente, o *theHarvester* é mais uma ferramenta *open-source* focada no mapeamento da superfície de ataque e no reconhecimento da infraestrutura de organizações, cuja principal aplicação se enquadra na identificação da superfície exposta de uma organização. Assim, este mapeamento funciona como ponto de partida para aferir o ciber-risco que incide sobre uma determinada entidade, nomeadamente no que concerne aos seus ativos expostos à *Internet* e a eventuais vulnerabilidades associadas. Desta forma, é utilizada no método de avaliação do ciber-risco organizacional de Lee e Shon [39]. Em particular, os autores recorrem a esta ferramenta para obter dados provenientes de fontes de informação sobre ativos como o *Censys* e o *Shodan*, cruzando-os, posteriormente, com informações sobre vulnerabilidades para identificar os ativos vulneráveis e potencialmente exploráveis, logo, agravadores do ciber-risco que impende sobre as organizações.

Contudo, sendo todas estas ferramentas de propósito geral, acabam por não ser das mais adequadas à compreensão do ciber-risco organizacional, por não terem sido concebidas com esta finalidade, mas sim numa perspetiva mais abrangente. Assim, conforme previamente referido, existem outras ferramentas desenvolvidas com o propósito específico de avaliar o ciber-risco de cada organização.

A Tabela 11 resume as ferramentas de propósito específico estudadas na literatura recolhida, associando-as às fontes de informação a partir das quais são recolhidos os dados e aos algoritmos de IA – como *eXtreme Gradient Boosting* (XGBoost), *Light Gradient-Boosting Machine* (LightGBM) e *Gradient Boosted Decision Trees* (GBDT), entre outros – utilizados para os tratar, com o objetivo de compreender o ciber-risco organizacional.

Ferramenta	Fontes	Algoritmos de IA	Referência
NiNSRAPM	*AbuseIPDB*, *FOFA, Shodan*	LightGBM, *Random Forest*, Regressão Logística, XGBoost	*[33]*
NiCSVAM	*AbuseIPDB*, *FOFA, Shodan*	LightGBM, *Random Forest*, Regressão Logística, XGBoost	*[34]*
ShoBeVODSDT	*BinaryEdge*, *Shodan*	Não Aplicável	*[31]*
Babenko *et al.*	*Censys*, CISA KEV, FIRST EPSS, NVD, *Shodan*, *WHOIS*	GBDT	*[30]*
Liu *et al.*	*DShield*, *PhishTank*, *SpamCop*, SURBL, VCDB	*Random Forest*	*[37]*
Sarabi *et al.*	VCDB	*Random Forest*	*[42]*

Tabela 11 – Ferramentas de Propósito Específico

Ao contrário das ferramentas de propósito geral anteriormente elencadas, as seis ferramentas da Tabela 11 já têm como propósito específico a avaliação do ciber-risco, sendo, por isso, mais adequadas para o efeito.

O *Non-Intrusive Network Security Risk Assessment Prediction Model* (NiNSRAPM), desenvolvido por Yang *et al.* [33] em 2022, contempla 25 indicadores obtidos através de fontes públicas de segurança da informação – como a existência de serviços vulneráveis, a presença de endereços IP em listas relacionadas com a divulgação de *spam*, *malware* ou *phishing* e a deteção de credenciais corporativas em fugas de dados – numa abordagem assente em aprendizagem automática para classificar o ciber-risco organizacional. Estes indicadores distribuem-se por nove dimensões, consideradas pelos autores como essenciais para avaliar o ciber-risco: segurança de redes, segurança de portas, segurança de DNS, segurança de *e-mail*, vulnerabilidades, segurança aplicacional, segurança de dados, ativos expostos e reputação de endereços IP. A partir destes atributos, o modelo avalia o ciber-risco numa escala de cinco níveis, combinando numa regressão logística as previsões de três algoritmos de classificação: LightGBM, *random forest* e XGBoost, obtendo cerca de 71% de exatidão e de pontuação F_1. Com isto, a solução proposta visa enfrentar o problema das seguradoras de cibersegurança em avaliar o ciber-risco de um cliente de forma não intrusiva, rápida e com baixo custo associado, sem precisar de autorização da organização em causa para aferir e classificar o respetivo ciber-risco.

Dos mesmos autores [34], o *Non-Intrusive Cyber Security Vulnerability Assessment Model* (NiCSVAM) segue uma abordagem semelhante, mas acaba por apresentar melhores resultados, sendo um desenvolvimento posterior da versão anterior, apresentado um ano mais tarde, em 2023. Esta ferramenta distingue-se do NiNSRAPM ao dividir os indicadores recolhidos a partir de fontes públicas em cinco dimensões em vez de nove, sendo elas: a reputação de endereços IP, a exposição de portas e/ou serviços, a segurança de certificados, a presença em fugas de dados e a segurança de aplicações *web*. De modo idêntico ao anterior, estes dados são processados por um modelo de aprendizagem automática assente nos mesmos três classificadores (LightGBM, *random forest* e XGBoost), resultando num valor enquadrado na mesma escala, desde A até E. Em comparação com a abordagem anterior, o NiCSVAM obtém resultados superiores em cerca de cinco pontos percentuais, com 76% de exatidão e 77% de pontuação F_1, demonstrando que um maior número de indicadores – como no NiNSRAPM – não se traduz, necessariamente, em melhores resultados preditivos. Igualmente, esta ferramenta visa servir o mesmo propósito da anterior, no que diz respeito à utilização por parte de seguradoras de cibersegurança para realizar uma avaliação preliminar do ciber-risco de potenciais clientes.

Da autoria de Daskevics e Nikiforova [31], o *Shodan and Binary Edge based Vulnerable Open Data Sources Detection Tool* (ShoBeVODSDT) integra, tal como o nome indica, fontes como o *Shodan* e o *BinaryEdge*[38] (descontinuado a 31 de março de 2025) para pesquisar ativos das organizações com o intuito de detetar bases de dados expostas e vulneráveis. O objetivo desta ferramenta passa por identificar bases de dados visíveis a partir de um ponto de vista externo às organizações, de maneira a avaliar se contêm dados sensíveis e, em última instância, se já foram comprometidas. Para este efeito, o ShoBeVODSDT assenta numa estratégia de funcionamento tripartida: primeiro, utiliza as APIs das duas fontes mencionadas para encontrar endereços IP de determinados serviços associados a sistemas de bases de dados; depois, tenta estabelecer uma ligação a cada um desses serviços, com o intuito de verificar se está, efetivamente, acessível externamente; em caso afirmativo, descarrega as informações constantes na base de dados, tendo em vista a aferição da sensibilidade dos dados e/ou um potencial comprometimento por *ransomware*. Embora esta ferramenta esteja limitada à avaliação do risco associado a bases de dados expostas, pode funcionar como um elemento relevante na avaliação global do ciber-risco que afeta uma organização, sendo este um possível contributo elencado pelos próprios autores.

Com uma finalidade mais abrangente, o artigo *Automated OSINT Techniques for Digital Asset Discovery and Cyber Risk Assessment* de Babenko *et al.* [30] incide precisamente no ponto de interseção entre a descoberta de ativos das organizações e a avaliação do ciber-risco associado. A ferramenta proposta recolhe dados públicos provenientes de múltiplas fontes – como registos DNS, informações sobre certificados SSL/TLS, entre outros – para identificar ativos expostos à *Internet*, integrando-os num modelo de aprendizagem automática constituído por GBDT. Na sua metodologia, os autores utilizam uma fórmula para calcular o ciber-risco avaliado para cada organização que considera não só a pontuação de cada vulnerabilidade – com base no seu CVSS e no respetivo EPSS – mas também a criticidade do ativo afetado – de acordo com a sensibilidade dos dados associados e o eventual impacto no negócio em caso de exploração – e o efeito de decaimento temporal devido ao tempo decorrido desde a divulgação da vulnerabilidade até ao momento da avaliação, bem como os potenciais efeitos de interação entre vulnerabilidades relacionadas para uma eventual propagação em cadeia. Esta abordagem resulta numa precisão superior a 90% nos casos avaliados e mostra-se, segundo os autores, extremamente escalável e eficaz.

[38] https://help.coalitioninc.com/hc/en-us/articles/34383910057371-BinaryEdge-Transition-FAQ

Existem, igualmente, soluções mais orientadas para a previsão da ocorrência de ciberincidentes que podem vir a afetar uma organização, sendo esta abordagem considerada pelos autores como uma forma mais concreta de aferição do ciber-risco [37], [42].

Em *Cloudy with a Chance of Breach: Forecasting Cyber Security Incidents,* Liu *et al.* [37] tentam prever a ocorrência de incidentes de cibersegurança tendo por base informações observáveis externamente às organizações. Para isto, dividem os dados extraídos em dois grupos: sintomas de má gestão – como configurações incorretas ou certificados inválidos – e sinais de atividade maliciosa – detetados pelas fontes de informação sobre ameaças previamente mencionadas. Assim, através de características extraídas a partir destas informações, como a duração e a frequência com que cada organização consta numa destas listas, a abordagem dos autores calcula, com 90% de precisão, a probabilidade de um futuro incidente numa organização, ao aplicar um modelo de *random forest*. Este estudo conclui que os sintomas de má gestão e os sinais de atividade maliciosa são fortes indicadores do ciber-risco organizacional, podendo servir para a previsão de ocorrência de ciberincidentes.

Por último, de Sarabi *et al.* [42], o artigo *Risky Business: Fine-Grained Data Breach Prediction using Business Profiles* propõe uma abordagem que estende a anterior, mas mais focada nas características de negócio de uma organização, com o intuito de prever eventuais fugas de dados futuras. Assim, ao cruzar atributos demográficos – como o setor e o número de colaboradores – com dados históricos de incidentes, os autores procuram não só antever a ocorrência de potenciais ciberincidentes, mas também identificar quais os tipos de ciberameaças (como erros humanos, ameaças internas ou ataques físicos, por exemplo) a que cada organização aparenta estar mais suscetível. Esta proposta tem resultados idênticos aos do artigo anterior, também através de um modelo de *random forest*. Tal como anteriormente, o objetivo desta ferramenta passa por contribuir para uma tomada de decisão mais proativa e informada em relação à cibersegurança das organizações.

Em suma, existem diversas ferramentas que recolhem, tratam e analisam informação a partir das fontes anteriormente descritas, tanto de propósito geral, como especificamente concebidas para avaliar o ciber-risco organizacional. A literatura evidencia que estas soluções incidem sobretudo na identificação de ativos expostos à *Internet*, vulnerabilidades conhecidas ou sinais externos de comprometimento, recorrendo, frequentemente, à agregação e à correlação de dados provenientes de múltiplas fontes públicas.

Todavia, estas ferramentas tendem a privilegiar indicadores associados à exposição da infraestrutura tecnológica, deixando menos exploradas outras formas de exposição digital que também podem constituir fatores relevantes de agravamento do ciber-risco organizacional, como é o caso da presença de credenciais corporativas em fugas de dados.

Adicionalmente, para além da identificação destes indicadores, importa também considerar a forma como a informação recolhida e analisada é posteriormente apresentada no âmbito do processo de OSINT.

2.2.3. Apresentação de Resultados

A última fase do processo de OSINT passa por apresentar os resultados obtidos, provenientes dos dados recolhidos das fontes públicas de informação, depois de devidamente tratados e analisados pelas ferramentas selecionadas [26].

Neste caso, as diferentes ferramentas de propósito específico elencadas anteriormente optam por abordagens distintas, assentes em escalas qualitativas ou quantitativas. A Tabela 12 sintetiza as várias formas de apresentação de resultados.

Ferramenta	Escala	Valor Mínimo	Valor Máximo	Referência
NiNSRAPM	Qualitativa	A (Excelente)	E (Muito Mau)	*[33]*
NiCSVAM	Qualitativa	A (Excelente)	E (Muito Mau)	*[34]*
ShoBeVODSDT	Qualitativa	0 (Falha de Conexão)	5 (Compro-metimento)	*[31]*
Babenko *et al.*	Qualitativa	Mínimo	Crítico	*[30]*
Liu *et al.*	Quantitativa	0	1	*[37]*
Sarabi *et al.*	Quantitativa	0	1	*[42]*

Tabela 12 – Formas de Apresentação de Resultados

Conforme ilustra a Tabela 12, as diferentes ferramentas seguem diferentes abordagens para a apresentação de resultados, ainda que, consoante a sua finalidade, possam utilizar escalas semelhantes.

O NiNSRAPM [33] e o NiCSVAM [34], enquanto ferramentas concebidas pelos mesmos autores – sendo a segunda um desenvolvimento posterior da primeira – apresentam o valor do ciber-risco percecionado para cada organização numa escala desde A até E. Nesta escala, o valor A (Excelente) significa ciber-risco mínimo, enquanto o valor E (Muito Mau) corresponde ao ciber-risco máximo. Esta escala tem como valores intermédios B (Bom), C (Médio) e D (Mau). Assim, sendo estas soluções focadas no mercado de seguros de cibersegurança, ambas procuram apresentar os resultados de forma simples, clara e inteligível, de forma a serem compreendidos, principalmente, pelas empresas seguradoras de cibersegurança para determinação dos montantes referentes aos seguros a disponibilizar.

Similarmente, o ShoBeVODSDT [31] utiliza uma escala qualitativa – apesar de numérica – de seis níveis para a apresentação de resultados. Nesta escala, as correspondências numéricas são as seguintes: 0 representa impossibilidade de ligação à base de dados; 1 corresponde a uma ligação bem-sucedida, mas sem possibilidade de recolher dados; 2 denota uma base de dados vazia; 3 significa recolha de dados, mas sem impacto para a organização (por exemplo, dados de sistema ou dados não sensíveis); 4 aponta para a extração de dados sensíveis e 5 mostra um comprometimento efetivo da base de dados em causa. Deste modo, no paradigma de compreensão do ciber-risco através da análise do estado das bases de dados corporativas expostas, esta escala traduz o grau de ciber-risco decorrente do nível de acesso possível à base de dados e da sensibilidade da informação obtida, refletindo o potencial impacto para a organização em causa.

Por sua vez, a abordagem de Babenko *et al.* [30] é a mais completa no que concerne à apresentação de resultados. Para esse efeito, a ferramenta desenvolvida contém uma camada de reporte que disponibiliza a informação obtida em múltiplos formatos, seja através de um *dashboard* operacional, seja por via de relatórios exportáveis, seja através de uma API para integração com outras soluções. No modelo dos autores, o nível de ciber-risco de cada organização é classificado numa escala qualitativa de cinco níveis: Mínimo, Baixo, Médio, Alto e Crítico, sendo acompanhado pelo grau de confiança atribuído ao mesmo e pelos fatores que contribuíram para esse resultado. Por isso, esta abordagem torna-se, simultaneamente, facilmente compreensível e explicável.

Finalmente, enquanto modelos que estimam a probabilidade de ocorrência de ciberincidentes futuros, as soluções de Liu *et al.* [37] e de Sarabi *et al.* [42] mostram os resultados sob a forma de uma probabilidade, ou seja, desde zero (0%) até um (100%). No primeiro modelo, o valor retornado traduz a probabilidade estimada de que um ciberincidente venha a afetar a organização em dois horizontes temporais distintos: a curto prazo, isto é, no espaço de um mês, ou a longo prazo, ou seja, num período até um ano. No segundo caso, o valor obtido corresponde à probabilidade aferida para cada organização quanto à ocorrência de um ciberincidente futuro, sem se cingir a um dado intervalo de tempo. Além disto, o segundo modelo apresenta também a probabilidade associada a cada tipo de potencial ameaça para a organização, isto é, quão suscetível cada organização está a ataques internos, ameaças externas, entre outros.

Em conclusão, as diferentes abordagens demonstram que a apresentação de resultados constitui uma parte essencial do processo de OSINT para a compreensão do ciber-risco. Para esse efeito, a maior parte das ferramentas opta por escalas qualitativas de tradução do grau/nível de ciber-risco, procurando uma forma facilmente compreensível, explicável e interpretável para apresentar os resultados.

2.2.4. Desafios e Limitações

A aplicação do processo de OSINT para a compreensão do ciber-risco organizacional tem, naturalmente, desafios e limitações [19]. Estas dificuldades podem surgir em qualquer uma das fases deste processo.

Em primeiro lugar, uma das principais limitações de qualquer processo de OSINT reside precisamente naquilo que lhe dá início: as fontes de informação [19]. A este respeito, os dois principais problemas são diametralmente opostos. Por um lado, a informação disponível publicamente sobre uma determinada organização pode ser escassa ou insuficiente, limitando, logo à partida, qualquer extração de conhecimento que dela dependa [33], [34]. Por outro lado, em sentido inverso, dado o crescente volume de dados disponível para a maior parte das entidades, pode tornar-se difícil a distinção entre a informação útil e o ruído cada vez mais omnipresente [19]. Em ambos os casos, a veracidade e a fiabilidade dos dados devem ser sempre contempladas, assim como a sua atualidade e pertinência [29]. Por tudo isto, a solução ideal tende a passar pela identificação de um conjunto restrito de fontes de informação que sejam, em simultâneo, suficientemente credíveis e abrangentes para os tipos de informação procurados, de modo a serem capazes de contribuir para a aferição do ciber-risco.

Em segundo lugar, passando para a fase de recolha de dados, também esta tem algumas dificuldades inerentes [19]. Tendo em conta a heterogeneidade das diversas fontes de informação, não só no que toca aos diferentes formatos em que a informação é disponibilizada e às múltiplas interfaces para obter os dados, mas também quanto à sua frequência de atualização e a outras restrições, a integração entre elas com o intuito de automatizar a recolha de dados nem sempre é vista como fácil [30]. Assim, enquanto a análise manual é considerada lenta e dificilmente escalável, o tratamento automático acrescenta uma complexidade não negligenciável, principalmente no momento de configuração inicial [19], [30], [40]. Para contornar isto, convém, caso seja possível, aproveitar soluções já existentes para a recolha de informação de múltiplas fontes, ainda que alguns autores considerem relativamente acentuada a curva de aprendizagem necessária para a utilização das mesmas [19], [40].

Transversalmente, existem implicações éticas, legais e sociais associadas à recolha de dados nos processos de OSINT [45]. A este respeito, podem ser levantadas questões quanto à legitimidade da utilização de dados públicos para fins diferentes dos quais foram originalmente destinados [19], bem como sobre as ferramentas que realizam *scans* periódicos para descobrirem ativos expostos [45]. Embora estas ferramentas sejam amplamente utilizadas, o seu enquadramento legal é, por vezes, questionado [45]. Por isso, podem ser preferíveis alternativas que não recorram a estas técnicas de mapeamento da superfície de ataque, mas que também sejam capazes de contribuir para a compreensão do ciber-risco organizacional, como é o caso da exploração e análise das fugas de dados com credenciais expostas, sendo estas um dos principais vetores atuais de intrusão nas organizações [44].

Em penúltimo lugar, no que concerne às ferramentas desenvolvidas com o propósito específico de avaliar o ciber-risco das organizações, a incompletude da visão obtida pelas mesmas também é um problema a considerar [38]. Como é visível, a maioria destas soluções acaba por se focar na deteção de vulnerabilidades em serviços expostos, assim como em potenciais falhas de configurações, embora existam mais vetores que podem ser considerados, como a exposição em fugas de dados, muitas vezes excluída da análise. Além disto, a visão obtida por estas ferramentas é, frequentemente, considerada limitada, na medida em que pode não ser capaz de captar uma mitigação ou contramedida interna para uma vulnerabilidade visível externamente [38]. A isto acresce ainda a inevitável existência de falsos positivos, isto é, deteções de vulnerabilidades exploráveis que, na verdade, não correspondem à realidade, por já se encontrarem desatualizadas ou por já terem sido solucionadas, por exemplo [33], [34].

Finalmente, especificamente para a análise do ciber-risco que afeta as organizações, nenhuma das soluções identificadas avalia um fator que, de acordo com o já referido DBIR, tem vindo a ganhar cada vez mais importância, por ser um vetor de ataque envolvido em 30% das fugas de dados identificadas no ano de 2025: o risco de terceiros [44]. Assim, a identificação, através de dados públicos, dos terceiros, fornecedores e prestadores de serviços das organizações – bem como a aferição do risco associado – também deve ser um fator a considerar, sendo esta uma lacuna transversal a todos os modelos de OSINT estudados para avaliar o ciber-risco organizacional.

Em síntese, embora existam soluções de OSINT para avaliar o ciber-risco das organizações, estas ainda continuam a ter falhas que podem ser solucionadas, como o foco extremamente orientado para as vulnerabilidades identificadas nos ativos expostos à *Internet* e a falta de visibilidade sobre os riscos de terceiros e da cadeia de fornecimento.

2.3. Riscos de Terceiros e da Cadeia de Fornecimento

Conforme referido anteriormente, os riscos de terceiros e da cadeia de fornecimento foram responsáveis por cerca de 30% das fugas de dados ocorridas em 2025 [44] e 10,6% dos ciberincidentes registados pela ENISA entre julho de 2024 e junho de 2025 [4], respetivamente. Por isso, estes vetores podem ser importantes na determinação do ciber-risco que afeta as organizações, pelo que também devem ser considerados para o efeito.

Antes de definir o conceito de risco da cadeia de fornecimento, convém clarificar o que se entende pela própria cadeia de fornecimento. O NIST define este termo como o "conjunto interligado de recursos e processos entre e ao longo de múltiplos níveis de organizações, sendo cada uma delas um adquirente, que se inicia com a aquisição de produtos e serviços e se estende ao longo do seu ciclo de vida" [46]. Por sua vez, segundo o mesmo organismo, o risco da cadeia de fornecimento é "o potencial de dano ou comprometimento que surge como resultado dos riscos de segurança provenientes de fornecedores, das suas cadeias de fornecimento e dos seus produtos ou serviços", incluindo "exposições, ameaças e vulnerabilidades associadas aos produtos e serviços que percorrem a cadeia de fornecimento, bem como as exposições, ameaças e vulnerabilidades à/da própria cadeia de fornecimento" [47].

Portanto, este conceito, conforme está definido, ao contemplar os "riscos de segurança provenientes de fornecedores" [47], engloba também a noção de risco de terceiros, enquanto fornecedores ou prestadores de serviços diretos. Efetivamente, na visão de Topping *et al.* [48] e de Weaver *et al*. [49], o risco de terceiros é entendido como sendo a camada mais direta, imediata e observável do risco da cadeia de fornecimento que, por sua vez, abrange um nível mais alargado e sistémico. Assim, segundo estes mesmos autores, uma estratégia de gestão de ciber-riscos que incida exclusivamente nos riscos de terceiros ou prestadores de serviços diretos revelar-se-á insuficiente, visto que ignora as restantes dependências entre as organizações da cadeia de fornecimento e o potencial de propagação do ciber-risco entre as mesmas que pode, em caso de comprometimento, afetar todas as organizações desta cadeia [48], [49].

Deste modo, para uma compreensão completa e abrangente dos riscos de terceiros e da cadeia de fornecimento, a literatura analisada entende que deve ser adotada uma abordagem holística, que abranja não só o risco direto proveniente de terceiros [50], [51], mas também o risco indireto resultante da propagação ao longo da cadeia de fornecimento [52], [53].

2.3.1. Avaliação do Risco de Terceiros

Na linha do que foi previamente referido, os autores estudados entendem que o primeiro passo para a compreensão do ciber-risco associado à cadeia de fornecimento passa pela avaliação do risco de terceiros, sendo esta a primeira camada da cadeia de fornecimento [48], [54].

Tradicionalmente, a metodologia mais utilizada pelas organizações para a avaliação do risco de terceiros residia no envio e preenchimento, por parte dos destinatários, isto é, das organizações terceiras a analisar, de questionários de segurança [55]. Estes questionários consistem numa lista de perguntas direcionadas ao terceiro, fornecedor ou prestador de serviços sobre as suas práticas de cibersegurança, que visam aferir a maturidade do mesmo nesta matéria para, posteriormente, serem tomadas decisões conforme as respostas fornecidas [55]. Contudo, como este método se baseia apenas em declarações do próprio terceiro, fornecedor ou prestador de serviços, nem sempre é considerado robusto, pelo que acaba por ser complementado – ou até substituído – por auditorias no local, ou seja, verificações técnicas pontuais capazes de comprovar as respostas fornecidas [56].

No entanto, esta abordagem tradicional traz limitações já amplamente reconhecidas pelas organizações. Por um lado, com o crescente desenvolvimento e a cada vez maior dependência da tecnologia, os questionários têm-se tornado progressivamente mais extensos e detalhados, aumentando o tempo e o esforço necessários para os elaborar e para lhes responder [55]. Por outro lado, as auditorias externas são dispendiosas, tanto em termos de tempo, como de recursos humanos, pelo que não se afiguram uma solução prática para todas as organizações [57]. Além disto, em ambos os casos, as avaliações não são contínuas, mas sim estáticas e respeitantes apenas ao momento temporal em que são efetuadas, pelo que, para manterem a sua utilidade e pertinência, evitando o risco de ficarem desatualizadas, requerem reavaliações periódicas, igualmente custosas [55], [58]. Acresce ainda que estas limitações são tanto maiores quanto maior for o número de terceiros abrangidos, logo, dificilmente escaláveis [57]. Por tudo isto, surge a necessidade de optar por outras soluções, preferencialmente capazes de avaliar o risco de terceiros de forma contínua e não intrusiva.

Na tentativa de dar resposta a todas estas limitações, surgiram os *ratings* de segurança que são, atualmente, a solução mais adotada pelas organizações para este efeito [59]. Um *rating* de segurança é um mecanismo que, de forma semelhante a um *rating* de crédito financeiro, atribui uma pontuação a cada organização conforme o risco avaliado. Assim, funciona como uma avaliação empírica e não intrusiva – por, na maior parte dos casos, se basear em dados disponíveis publicamente – que visa avaliar o ciber-risco intrínseco de uma organização e, por extensão, o risco que prolifera para os seus clientes e parceiros [59].

Atualmente, existe uma vasta panóplia de soluções de *ratings* de segurança. Apesar desta diversidade, de acordo com Keskin *et al.* [59], estas soluções tendem a contemplar seis vetores comuns para aferir o ciber-risco organizacional. Na ótica destes autores, esses vetores são os seguintes: (1) comportamento dos *endpoints*, (2) configurações de rede, (3) *software* e serviços, (4) domínios *web*, (5) informação sobre a organização e (6) histórico de fugas de dados.

Em primeiro lugar, a avaliação do comportamento dos *endpoints* consiste, na ótica dos autores, em procurar por eventuais evidências de comprometimento dos computadores e dos servidores das organizações através da presença dos mesmos em listas de distribuição de *spam* ou *malware*, como as da Tabela 9. Assim, caso um ativo organizacional esteja presente numa destas listas, tal será indicativo de um eventual comprometimento, logo, agravador do ciber-risco da organização [59].

Em segundo lugar, a análise das configurações de rede visa compará-las com as melhores práticas para os dispositivos de rede, tendo em vista a identificação de erros que podem comprometer a cibersegurança das organizações. Por exemplo, este vetor deve contemplar a deteção de portas abertas, certificados desatualizados e protocolos descontinuados, sendo todos estes fatores de risco quando indevidamente configurados, segundo os mesmos autores [59].

Em terceiro lugar, a componente de *software* e serviços engloba a identificação das versões dos sistemas expostos e o mapeamento das mesmas perante fontes de informação sobre vulnerabilidades como as da Tabela 6, enquanto informação que também contribui para avaliar o ciber-risco das organizações. Segundo Keskin *et al.* [59], ainda nesta categoria pode ser incluída a presença de determinados mecanismos de segurança *web*, como, por exemplo, a existência de determinados cabeçalhos *HyperText Transfer Protocol* (HTTP) e a presença de mecanismos como HTTP *Strict Transport Security* (HSTS), entre outros.

A par disto, o vetor relacionado com os domínios *web* abrange, para além das vulnerabilidades associadas ao domínio da organização, toda a informação proveniente de registos DNS, respostas a pedidos *Internet Control Message Protocol* (ICMP), etc., dado que, tal como anteriormente, os autores consideram que todos estes parâmetros podem permitir a avaliação do ciber-risco organizacional [59].

Em penúltimo lugar, a informação sobre a organização, conforme já referido por Sarabi *et al.* [42], também pode ser útil para o mesmo efeito. A este respeito, dados como a dimensão, o estado financeiro e o setor de atividade, juntamente com endereços IP e endereços de *e-mail*, por exemplo, são vistos por Keskin *et al.* [59] como elementos que podem contribuir para facilitar ataques de engenharia social, expondo a organização em causa a um maior ciber-risco, constituindo, por isso, fatores que devem ser considerados na avaliação.

Por último, em consonância com o exposto por Liu *et al.* [37], o histórico de fugas de dados respeitantes a uma organização pode contribuir para a aferição do respetivo ciber-risco. De acordo com Keskin *et al.* [59], estas informações permitem avaliar a capacidade de uma organização para gerir e mitigar ciber-riscos, sendo um dos indicadores com maior capacidade para prever futuros incidentes que, no caso de resposta inadequada, podem comprometer não só a própria organização, mas também as entidades com as quais esta está direta ou indiretamente relacionada através da cadeia de fornecimento.

Enquanto exemplos de abordagens que operacionalizam estes seis vetores para avaliar o ciber-risco organizacional, Keskin *et al.* [59] destacam nove soluções comerciais de avaliação do risco de terceiros.

A primeira é a *BitSight*[39], sendo esta uma das empresas de *ratings* de cibersegurança mais reputadas no mercado. Na sua abordagem, são combinados os dados provenientes de mais de 120 fontes públicas de informação num modelo proprietário, para se obter como resultado um valor entre 250 e 900, que representa o ciber-risco percecionado para cada organização [59]. Estas classificações podem ser consultadas através do serviço comercializado pela *BitSight*, que permite visualizar os *ratings* em *dashboards* dedicados na plataforma ou integrá-los em sistemas externos por via de uma API.

A *ComplyScore*[40] é uma outra solução de *ratings* de cibersegurança, mas com particular foco na conformidade das organizações. Esta plataforma procura gerir os riscos de terceiros e da cadeia de fornecimento através do mapeamento das práticas e dos controlos de segurança de cada organização com as diretivas, *frameworks*, normas e regulamentos aplicáveis, realizando uma avaliação do ciber-risco guiada pela conformidade [59]. O acesso aos resultados desta avaliação está condicionado aos utilizadores desta plataforma paga, que visa apoiar as organizações nas atividades de monitorização de conformidade e de avaliação do risco de terceiros, fornecedores e prestadores de serviços.

A *Fair Isaac Corporation* (FICO)[41] é uma empresa de análise de dados focada na elaboração de *ratings* de crédito, complementados por *ratings* de ciber-risco, entre outras informações. Para esse efeito, a FICO recorre a um modelo de aprendizagem automática proprietário com o intuito de determinar o perfil de risco de cada organização, tendo por base dados históricos associados a múltiplos indicadores capazes de aferir a postura dos respetivos sistemas, redes, *software* e serviços, bem como eventuais comprometimentos ocorridos no passado [59]. Neste caso, as classificações são fornecidas no âmbito das soluções comerciais da FICO para avaliação do risco, sendo, sobretudo, utilizadas por organizações e instituições financeiras que recorrem a estes serviços para apoiar processos formais de análise e tomada de decisão.

[39] https://www.bitsight.com/
[40] https://www.atlassystems.com/complyscore
[41] https://www.fico.com/

A *Interos*[42] é mais uma solução de avaliação do risco de terceiros, também apoiada por modelos de IA, mas numa perspetiva alargada da cadeia de fornecimento, indo para além dos terceiros, fornecedores e prestadores de serviços diretos. Contrariamente às outras abordagens referidas, a *Interos* não atribui uma pontuação de ciber-risco a cada organização, mas disponibiliza indicadores de possíveis vulnerabilidades existentes em cada entidade integrante da cadeia de fornecimento. Com isto, pretende-se que estes dados sejam mais informativos para os utilizadores do que um simples indicador do grau de ciber-risco associado a uma organização, sem uma explicação associada [59]. A informação produzida por este sistema é disponibilizada aos clientes da *Interos* através das suas ferramentas de inteligência da cadeia de fornecimento.

A *Recorded Future*[43]*,* enquanto solução especializada na recolha de informação sobre ciberameaças, monitoriza continuamente todas as organizações registadas na sua base de dados, através dos seus módulos de inteligência de terceiros e de vulnerabilidades. Esta solução recolhe e analisa dados públicos disponíveis na *surface web* e na *dark web* com algoritmos proprietários de aprendizagem automática e NLP, tendo em vista a identificação de novas vulnerabilidades e de evidências de comprometimento das organizações, apresentando os resultados numa escala de zero a cem [59]. Os indicadores resultantes desta monitorização contínua podem ser explorados nos módulos de inteligência da plataforma da *Recorded Future*, assim como incorporados noutras soluções, através das interfaces de integração pagas disponibilizadas pelo serviço.

A *RiskRecon*[44] (adquirida em 2019 pela *Mastercard*[45]) avalia de forma contínua os sistemas de cada organização contra 39 critérios, de maneira a modelar o respetivo ciber-risco com base em informações públicas. Também nesta solução, os dados recolhidos alimentam um modelo de IA, resultando numa pontuação de um a dez que tem em conta a priorização de determinados vetores de risco conforme o valor do ativo afetado [59]. Estas pontuações são apresentadas no serviço de monitorização da *RiskRecon*, permitindo acompanhar a evolução do ciber-risco das organizações analisadas ao longo do tempo.

[42] https://www.interos.ai/
[43] https://www.recordedfuture.com/
[44] https://www.riskrecon.com/
[45] https://www.mastercard.com/

A *Black Kite*[46] é um sistema de *ratings* de cibersegurança orientado pelo possível impacto financeiro que pode surgir devido a um ciberataque num terceiro, fornecedor ou prestador de serviços. Esta abordagem contempla três principais componentes para avaliar o ciber-risco organizacional: (1) o *rating* de cibersegurança aferido de forma similar às outras ferramentas, (2) verificações de conformidade e (3) o valor provável de perda financeira, determinado por uma metodologia de quantificação de ciber-riscos. O resultado da combinação destes vetores é traduzido numa escala desde A até F [59], disponibilizado na plataforma comercializada por esta solução.

Por sua vez, a *Panorays*[47] combina a abordagem das soluções de *ratings* com a metodologia tradicional de avaliação do risco de terceiros através de questionários. Assim, esta solução permite, por um lado, submeter questionários personalizados – por exemplo, com base nas diretivas, normas e regulamentos aplicáveis – e, por outro lado, observar a classificação do ciber-risco percecionado através de fontes públicas [59]. As classificações resultantes deste cruzamento de informações são fornecidas aos utilizadores através desta ferramenta paga.

Finalmente, a *SecurityScorecard*[48] avalia a postura de cibersegurança de cada organização tendo por base vários indicadores públicos, tais como informação sobre ameaças, vulnerabilidades e ativos expostos à *Internet*, etc. Estes indicadores são mapeados em diversas categorias, como segurança de aplicações *web*, segurança de redes, DNS, fugas de dados, reputação, suscetibilidade a engenharia social, menções em fóruns de cibercriminosos, tempo de resposta e aplicação de *patches*, resultando numa avaliação de zero a cem e, correspondentemente, de A até F [59]. Tal como nos outros casos, também a *SecurityScorecard* é uma solução comercial, frequentemente integrada em processos de gestão de risco de terceiros, tanto por via da sua plataforma, quanto através da respetiva API.

Em resumo, embora estas soluções de avaliação do risco de terceiros constituam uma abordagem robusta e abrangente quanto ao que se propõem realizar, todas elas são soluções comerciais, que só disponibilizam os seus resultados para as organizações que adquiram o respetivo licenciamento. Por isso, surge a oportunidade de adotar uma metodologia semelhante, mas de forma mais facilmente acessível às organizações, conforme apontam alguns autores na literatura analisada [57].

[46] https://blackkite.com/
[47] https://panorays.com/
[48] https://securityscorecard.com/

Efetivamente, Asili e Bahtiyar [57] referem que as soluções de *ratings* de cibersegurança são, frequentemente, demasiado dispendiosas, em particular para as pequenas e médias empresas (PME), cujos recursos são particularmente limitados. Por isso mesmo, estes autores propõem uma nova metodologia para avaliar o risco de terceiros, especialmente direcionada para as PME, designada *SME-SHIELD*. Esta metodologia assenta em quatro principais etapas. Em primeiro lugar, modela os terceiros como nós num grafo não dirigido, sendo cada nó constituído por um vetor de atributos de risco, nos quais se incluem, entre outros, uma classificação da indústria na qual esse terceiro se insere, um índice de exposição calculado a nível regional, um indicador do nível de conformidade do terceiro e a frequência de ocorrência de ciberincidentes no seu histórico. Em segundo lugar, o *SME-SHIELD* recolhe dados de fontes públicas, através de um processo de OSINT, criando um outro vetor de atributos. Entre as informações recolhidas, incluem-se, nomeadamente, a exposição das versões dos serviços e sistemas do terceiro em bases de dados de vulnerabilidades, a presença do mesmo em plataformas de fugas de dados na *dark web*, a frequência de publicações nos meios de comunicação social sobre fugas de dados e alguns indicadores de reputação externa, como da *BitSight* ou da *SecurityScorecard*, mas apenas quando estas informações se encontram acessíveis publicamente, de maneira a não incorrer nos mesmos custos inicialmente vistos como problemáticos. Com isto, estes dois vetores de atributos são combinados num só, que é utilizado como valor de entrada num conjunto heterogéneo de modelos de aprendizagem automática, constituído por cinco algoritmos – LightGBM, *Naïve Bayes*, *random forest*, regressão logística e XGBoost – cujos resultados são ponderados através da atribuição de diferentes pesos a cada um. Finalmente, o valor retornado por este conjunto de modelos é enquadrado numa escala categórica de risco, sendo classificado como baixo, médio ou alto. Através desta abordagem, os autores alcançaram uma exatidão de 84% e uma precisão de 86% nos testes efetuados.

Deste modo, seja através dos *ratings* comerciais de cibersegurança, seja por via desta nova arquitetura proposta, existem soluções capazes de avaliar o risco de terceiros de forma não intrusiva. No entanto, esta é apenas a primeira camada da cadeia de fornecimento.

2.3.2. Propagação do Risco da Cadeia de Fornecimento

De acordo com as perspetivas já explanadas de alguns autores, a avaliação do risco da cadeia de fornecimento além dos terceiros diretos – ao incluir também os fornecedores e prestadores de serviços indiretos ao longo da cadeia e as relações de interdependência entre as organizações – permite uma visão mais alargada, mas igualmente necessária, para a compreensão do ciber-risco organizacional [48], [49].

A razão para esta relevância dada à cadeia de fornecimento deve-se àquilo que Ghadge *et al.* [51] denominam "efeito cascata". Este conceito traduz a ideia de que, sendo a cadeia de fornecimento uma rede formada por diferentes organizações interconectadas entre si, o comprometimento ou a falha numa única entidade pode resultar num impacto com potencial de propagação estrutural e transversal a toda a cadeia de fornecimento. Com isto, os autores entendem que a segurança de uma organização já não depende apenas da gestão do ciber-risco que incide diretamente sobre ela, mesmo que também sejam considerados os terceiros ou prestadores de serviços diretos, mas que toda a cadeia de fornecimento deve ser tida em conta, enquanto potencial fator de risco sistémico que pode acabar por afetar todas as organizações que dela fazem parte.

Nesta medida, Ghadge *et al.* [51] definem três tipos de consequências de um ciberataque ou incidente de cibersegurança: as consequências primárias, secundárias e terciárias. As consequências primárias consistem no impacto direto do incidente que afeta a organização comprometida enquanto principal alvo. São exemplos destas consequências a interrupção das operações dessa organização e a perda de produtividade associada, derivada de um ataque direto. As consequências secundárias englobam a propagação do impacto de um incidente para a restante cadeia de fornecimento, para além da principal entidade visada pelas consequências primárias. De acordo com os autores, as relações entre as organizações – ao facilitarem a partilha de informação e possibilitarem a colaboração necessária para o desenrolar das respetivas atividades – também permitem, por exemplo, a transmissão de *malware* e a exposição ou fuga de dados sensíveis de entidades que não a diretamente visada, o que resulta em danos que vão para além do alvo original. Finalmente, as consequências terciárias traduzem o impacto na sociedade em geral, quando este extravasa a própria cadeia de fornecimento, podendo chegar, em casos mais extremos, a alcançar entidades que nem dela fazem parte. Considerando tudo isto, os autores defendem a importância de um modelo unificado de gestão de ciber-riscos na cadeia de fornecimento.

Para clarificar a forma através da qual os ciber-riscos se podem propagar entre as organizações, resultando nas consequências secundárias e terciárias referidas anteriormente, Filho *et al.* [60] consideram dois meios: o virtual e o físico. No seu entender, o meio virtual – através da informação partilhada – é o principal veículo de propagação dos ciberataques e do ciber-risco. Segundo argumentam, por exemplo, em casos de *phishing* e *malware*, a troca de informações entre organizações pode permitir a disseminação de um incidente e das suas consequências. A par disto, o ciber-risco também se pode propagar através do meio físico, como o *hardware*. Este caso – embora seja considerado pelos autores como sendo mais específico e direcionado – tem como exemplos a falsificação ou a manipulação indevida de dispositivos físicos, antes de serem entregues ou partilhados entre as organizações. Assim, ambas as situações constituem veículos de propagação do ciber-risco na cadeia de fornecimento.

Noutra perspetiva, Deane *et al.* [61] propõem uma taxonomia para compreender a propagação do ciber-risco ao longo da cadeia de fornecimento. Esta taxonomia considera dois eixos: a relação entre as organizações e a topologia da rede na qual se inserem.

No primeiro eixo, Deane *et al*. [61] entendem que existem três tipos de relações possíveis entre uma organização e os seus fornecedores, que traduzem a forma como uma organização pode ser afetada: parceiro-vetor, parceiro-guarda e parceiro-ator. Na relação parceiro-vetor, o fornecedor é o vetor de transmissão do ataque/incidente, sendo comprometido como um meio para atingir um fim. Na relação parceiro-guarda, o fornecedor limita-se a armazenar dados da organização em análise, pelo que o seu comprometimento e uma eventual violação ou fuga desses dados impacta também a organização. Finalmente, na relação parceiro-ator, o fornecedor é o próprio agente malicioso que atua com intenção de atacar a organização, procurando tirar proveito das relações existentes.

Ortogonalmente, o eixo relativo à topologia de rede também considera três casos possíveis: uma rede sequencial, centralizada ou recíproca. Na rede sequencial, tal como o nome indica, a relação entre as diferentes entidades é uma sequência, isto é, uma primeira entidade presta serviços a uma segunda, que, por sua vez, presta serviços a uma terceira e assim sucessivamente. Na rede centralizada, existe uma entidade central, que se relaciona com muitas outras organizações da cadeia de fornecimento, pelo que é um ponto crítico de falha em caso de comprometimento. Por último, na rede recíproca, existem ligações entre todas as entidades da cadeia, o que, na perspetiva dos autores, se torna num fator agravador do ciber-risco organizacional [61].

Neste contexto de propagação do ciber-risco entre as organizações interligadas ao longo da cadeia de fornecimento, alguns trabalhos estudados na literatura têm procurado analisar o fenómeno através de abordagens de modelação quantitativa. Em particular, têm sido propostos modelos matemáticos que permitem estimar a probabilidade de comprometimento entre organizações interligadas e averiguar de que forma é que um ciberincidente numa entidade se pode propagar a outras devido às relações existentes na cadeia de fornecimento. Embora estas abordagens não incorporem diretamente as tipologias ou os mecanismos conceptuais discutidos anteriormente, procuram quantificar o fenómeno de propagação do ciber-risco ao longo da cadeia de fornecimento. Entre estes contributos, destacam-se os modelos propostos por Li e Xu (em 2021) [53], por Dash *et al.* (em 2024) [52] e por Mondal e Singh (em 2025) [62].

A proposta de Li e Xu [53] modela a propagação do ciber-risco na cadeia de fornecimento considerando diversos acontecimentos e as respetivas probabilidades. Em primeiro lugar, os autores consideram a existência de uma probabilidade de propagação (q), que representa a possibilidade de um ataque ou incidente ocorrido num fornecedor acabar por afetar uma outra organização. De acordo com os autores, este valor é assumido como sendo uma constante para cada organização, ou seja, comum a todos os seus fornecedores. Em segundo lugar, os autores distinguem dois tipos de comprometimento. Por um lado, existe a probabilidade de comprometimento direto da organização (representada por p_a), que corresponde a um ataque direcionado à própria organização e que resulta do nível do ciber-risco intrínseco da organização. Por outro lado, considera-se a probabilidade de comprometimento de cada fornecedor i (traduzida por p_{b_i}), isto é, a probabilidade de esse fornecedor ser comprometido, também determinada pelo respetivo grau de ciber-risco.

No modelo desenvolvido, o cálculo da probabilidade de comprometimento indireto de uma organização através de um fornecedor i (designada por P_{a_i}) consiste em multiplicar o valor correspondente à probabilidade de propagação (q) do ciber-risco entre esse fornecedor e a organização pela probabilidade de o próprio fornecedor ser comprometido (p_{b_i}). Ou seja, a probabilidade de comprometimento indireto de uma organização através do seu fornecedor i pode ser obtida pelo cálculo da expressão traduzida na Equação (1) [53].

$$P_{a_i} = q p_{b_i} \tag{1}$$

Com isto, é possível calcular a probabilidade de uma organização ser comprometida (P_a). Para tal, os autores recorrem ao acontecimento complementar, considerando que uma organização só não é comprometida no caso da conjugação simultânea de dois acontecimentos: (1) se não for comprometida diretamente e (2) se não for comprometida indiretamente por nenhum dos seus n fornecedores. Assim, a probabilidade de uma organização não ser comprometida corresponde ao produto entre as probabilidades destes dois acontecimentos. Consequentemente, a probabilidade total de comprometimento da organização corresponde ao evento complementar deste, conforme expresso na Equação (2) [53].

$$P_a = 1 - (1 - p_a) \prod_{i=1}^{n} \left(1 - P_{a_i}\right) \quad (2)$$

Apesar de este artigo propor uma metodologia para o cálculo da propagação do ciber-risco entre organizações – por via da probabilidade de comprometimento –, os autores só aplicam o modelo a cenários em que a propagação ocorre entre dois nós vizinhos da cadeia de fornecimento, sem intermediários, isto é, apenas entre a organização e os seus fornecedores diretos. Por isso, este modelo limita-se a considerar uma propagação do ciber-risco de um só estágio único, não contemplando situações em que um incidente possa atravessar múltiplos níveis da cadeia de fornecimento. Além disto, também é assumido pelos autores que a probabilidade de propagação (q) é idêntica para todos os fornecedores de uma organização, sem diferenciar o grau de dependência da organização relativamente a diferentes fornecedores, o que pode consubstanciar uma limitação [53].

Assim, o modelo de Dash *et al.* [52] estende o trabalho anterior ao generalizar a modelação da propagação do ciber-risco para dois estágios ao longo da cadeia de fornecimento. Neste caso, para além da possibilidade de propagação direta entre duas organizações interligadas, os autores consideram também cenários em que o comprometimento de uma entidade pode ser propagado para uma organização intermédia e, posteriormente, desta para uma terceira. Estes autores replicam a mesma lógica probabilística apresentada anteriormente, continuando a distinguir o comprometimento direto do comprometimento indireto. No entanto, este modelo passa a considerar explicitamente situações em que a propagação do ciber-risco ocorre através de uma entidade intermédia, permitindo representar cenários em que o risco se propaga ao longo de dois níveis da cadeia de fornecimento.

Sendo p_a a probabilidade de comprometimento direto da organização em análise, p_b e p_c as probabilidades de comprometimento das organizações interligadas, enquanto q e r representam as respetivas probabilidades de propagação do ciber-risco entre essas entidades, a Equação (3) concretiza a probabilidade total de comprometimento (P_a), de acordo com este modelo [52].

$$P_a = 1 - (1 - p_a)(1 - qp_b)(1 - rp_c) \quad (3)$$

Embora o artigo original considere apenas uma entidade em cada nível da cadeia de fornecimento, a formulação proposta pode ser estendida a cenários mais complexos, com múltiplas entidades interligadas em cada nível. Nesse caso, deve ser mantida a mesma lógica probabilística assente no acontecimento complementar, mas alargando o cálculo de modo a incluir, no produto das probabilidades de não comprometimento indireto, todas as entidades relevantes e as respetivas vias de propagação. Com isto, a abordagem de Dash *et al.* [52] acaba por ser mais completa e abrangente do que a anterior, ainda que continue a assumir uma estrutura de propagação linear ao longo da cadeia de fornecimento.

Por último, Mondal e Singh [62] propõem uma abordagem mais robusta à propagação do ciber-risco na cadeia de fornecimento, indo para além da topologia linear dos estudos anteriores. Para isso, os autores recorrem a uma matriz de adjacências que mapeia as ligações entre todos os nós da rede, considerando a probabilidade de propagação do risco entre cada par de entidades. Nesta matriz, cada ligação pode assumir um peso distinto, refletindo diferentes graus de interdependência entre os membros da rede, o que permite representar de forma mais realista a intensidade com que o ciber-risco pode ser transmitido entre organizações. Posteriormente, são realizadas simulações de Monte Carlo para estudar a propagação do ciber-risco entre as entidades. Deste modo, os autores mostram que as entidades com várias relações concentram nelas próprias o maior risco da cadeia de fornecimento, criando potenciais impactos em cascata por serem um ponto crítico de falha.

Em suma, a literatura entende que o ciber-risco de uma organização não pode ser visto como um elemento isolado da cadeia de fornecimento, devendo ser tratado de forma integrada e sistémica, dado o seu potencial de propagação em cascata. Para isso, é essencial que o modelo adotado para o analisar seja capaz de captar, de forma adequada, as relações de interdependência entre as organizações e o seu papel na propagação do ciber-risco ao longo da cadeia de fornecimento.

2.3.3. Desafios e Limitações

Todas as soluções apresentadas – tanto a nível da avaliação do risco de terceiros, quanto para a modelação da propagação do ciber-risco ao longo da cadeia de fornecimento – apresentam os seus desafios e limitações [49].

Em primeiro lugar, relativamente às abordagens tradicionais de avaliação do risco de terceiros, como os questionários de segurança e as auditorias externas, alguns autores entendem que estas tendem a revelar-se incompletas e ineficazes. Isto porque, na ótica de Viega e Michael [55], as respostas aos primeiros podem não refletir a realidade operacional das organizações e a realização das segundas é altamente dispendiosa em termos de recursos. A isto acresce ainda o facto de ambas as metodologias só serem capazes de captar a realidade observável num único instante no tempo, isto é, quando são efetivamente concretizadas, sendo, por isso, estáticas, logo, consideradas insuficientes por diversos autores [56]. Nesta medida, surge a necessidade de recorrer a uma abordagem capaz de avaliar continuamente os terceiros relativamente ao seu grau de ciber-risco.

Em segundo lugar, os *ratings* de segurança, enquanto solução proposta para os problemas anteriores, acabam por sofrer de uma limitação semelhante: a falta de consistência entre avaliações. De acordo com o estudo de Keskin *et al*. [59], avaliações relativas à mesma organização, mas realizadas por diferentes soluções de *ratings*, nem sempre convergem para o mesmo resultado, o que suscita dúvidas sobre a sua fiabilidade. A par disto, estas abordagens só têm capacidade para observar a pegada digital externa de cada organização, o que, efetivamente, pode deixar escapar controlos internos de mitigação das eventuais falhas identificadas [63]. Por isso, esta ainda não é considerada a solução ideal para a avaliação do risco de terceiros.

A tudo isto, acresce ainda a dificuldade de uma organização identificar de forma correta e completa os seus terceiros, fornecedores e prestadores de serviços, tendo a necessidade de distinguir aqueles que são relevantes para a segurança da sua informação daqueles que não o são. Pinto *et al.* [58] realçam a existência deste problema no que concerne à volatilidade e à diversidade de ambientes *cloud* nos quais uma organização armazena ou trata os seus dados, o que, no seu entender, dificulta o levantamento efetuado pelas soluções de *ratings* apresentadas, mas a mesma questão pode ser suscitada para o mapeamento e a identificação dos parceiros propriamente ditos, dada a vasta panóplia de terceiros, fornecedores e prestadores de serviços que uma organização pode ter.

No que concerne ao nível mais estendido da cadeia de fornecimento, ou seja, além dos fornecedores diretos, alguns autores apontam como desafio crítico a incapacidade de uma organização ver para além desse mesmo primeiro nível, isto é, identificar as organizações subcontratadas e as dependências ocultas associadas [48], [49]. Na perspetiva destes autores, a falta de visibilidade sobre as entidades a montante na cadeia de fornecimento é uma das maiores dificuldades subjacentes à avaliação do risco associado. Assim, torna-se importante a existência de uma solução capaz de dotar as organizações de um conhecimento mais abrangente da cadeia de fornecimento na qual estão inseridas.

Finalmente, outros autores consideram também que, para além desta principal limitação, existe ainda a dificuldade de quantificar, medir ou avaliar a capacidade de propagação do ciber-risco ao longo da cadeia de fornecimento, apesar dos modelos matemáticos que a procuram solucionar [51], [62]. Sendo a quantificação de ciber-riscos um problema já intrinsecamente complexo, mas necessário para contribuir no sentido de auxiliar as tomadas de decisão [64], este problema assume ainda maiores proporções quando se trata de avaliar matematicamente a propagação do ciber-risco ao longo da cadeia de fornecimento [51], [62]. Assim, embora seja amplamente reconhecido que o ciber-risco é sistémico e que é passível de ser propagado em cascata, ainda não é clara na literatura a forma mais adequada para modelar esta propagação.

Em conclusão, a avaliação dos riscos de terceiros e da cadeia de fornecimento ainda se encontra numa fase de evolução. Embora já tenha sido percorrido um caminho – desde os questionários e as auditorias para os *ratings* de segurança e para modelos matemáticos –, ainda subsistem problemas por endereçar, desde logo no que diz respeito à correta identificação das entidades integrantes da cadeia de fornecimento e das relações entre elas.

2.4. Considerações Finais

Concluído o levantamento do estado da arte, cumpre tecer algumas considerações finais em resposta às questões de investigação previamente suscitadas, tendo por base a informação recolhida e o conhecimento adquirido nesta revisão de literatura.

Relativamente à pergunta de investigação principal, sinteticamente, a informação pública pode ser analisada para compreender o ciber-risco associado a uma entidade e às suas relações com terceiros através de um processo de OSINT estruturado, desde a definição das fontes de informação até à apresentação de resultados, passando pela recolha, tratamento e análise dos dados. A informação recolhida para este efeito pode abranger diversas categorias, como vulnerabilidades, ativos, ataques e ameaças, tanto da própria organização, quanto de entidades externas. No entanto, embora se reconheça que a análise do ciber-risco organizacional deva englobar a própria organização e os seus terceiros, fornecedores e prestadores de serviços, idealmente ao longo de toda a cadeia de fornecimento, muitas das soluções estudadas concentram-se num só eixo, isto é, ou na própria organização, ou nos terceiros. Assim, raramente são captadas dependências indiretas, pelo que falta uma abordagem holística unificada a todo o panorama de ciber-riscos, que disponha de uma maior visibilidade. Esta limitação revela uma lacuna estrutural na literatura, que tende a tratar o ciber-risco organizacional como um fenómeno isolado, em vez de um fenómeno sistémico assente em relações de interdependência entre entidades. As seguintes respostas às quatro questões de investigação secundárias detalham estas considerações.

Em primeiro lugar, quanto ao estado da arte relativamente à utilização de OSINT para compreender o ciber-risco de uma organização, a abordagem predominante nas soluções analisadas passa por considerar diferentes fatores para avaliar o ciber-risco organizacional. Entre estes, a vasta maioria da literatura analisada tende a combinar os ativos de uma organização expostos à *Internet* com as vulnerabilidades que lhes estão associadas, cruzando estas duas informações para obter como resultado uma pontuação correspondente ao ciber-risco avaliado. Esta predominância evidencia uma forte orientação para indicadores técnicos facilmente observáveis, o que acaba por restringir a compreensão do ciber-risco a um conjunto relativamente limitado de fatores. Ainda assim, dada a falta de soluções que explorem outras alternativas igualmente capazes de aferir o grau de ciber-risco, existe abertura para serem identificados diferentes fatores tendo em vista a prossecução do mesmo objetivo, como é o caso, por exemplo, da exposição de credenciais corporativas em fugas de dados. Este tipo de informação, apesar de ser amplamente associado à materialização de ciberataques em relatórios de referência na indústria, permanece ainda pouco explorado nas abordagens académicas analisadas.

Em segundo lugar, a informação necessária para compreender o ciber-risco de uma organização tem-se cingido, essencialmente, ao cruzamento dos ativos expostos à *Internet* com as suas vulnerabilidades, ao histórico de ataques contra a organização e a informações sobre ameaças. Assim, embora seja reconhecido como necessário, as soluções estudadas acabam por não considerar – em conjunto com esta informação – dados referentes aos terceiros, fornecedores e prestadores de serviços, igualmente essenciais para compreender o ciber-risco organizacional. Consequentemente, a análise tende a permanecer limitada à entidade individual, não refletindo adequadamente as relações que caracterizam as cadeias de fornecimento atuais.

Em terceiro lugar, as ferramentas existentes para recolher e tratar esta informação podem ser de propósito geral ou específico para o efeito. As ferramentas de propósito geral, embora recolham, combinem e correlacionem dados provenientes de diversas fontes de informação, acabam por não ser tão adequadas para compreender o ciber-risco organizacional, visto que não foram projetadas com esta finalidade. Na prática, estas ferramentas privilegiam a agregação de informação e a análise exploratória de dados, mas não a modelação do ciber-risco. Quanto às ferramentas de propósito específico, das quais são exemplos o ShoBeVODSDT [31] e as propostas de Babenko *et al.* [30], Liu *et al.* [37] e Sarabi *et al.* [42], nenhuma destas trata a exposição digital aferida por via das credenciais detetadas em fugas de dados como um elemento integrante da avaliação do ciber-risco, pese embora este seja um fator relevante para a consumação de ciberataques, conforme apontam os mais recentes relatórios [44].

Por último, os principais desafios e limitações na utilização de OSINT para compreender o ciber-risco de uma organização surgem tanto na vertente do ciber-risco intrínseco, quanto nos riscos de terceiros e da cadeia de fornecimento. Para além dos inevitáveis problemas associados à existência de falsos positivos e à dependência da qualidade dos dados, as abordagens que tentam aferir o ciber-risco organizacional através de um processo de OSINT tendem a não captar o risco de terceiros e da cadeia de fornecimento, enquanto as soluções que o fazem acabam por se focar exclusivamente neste aspeto. No caso particular destas últimas soluções, muitas padecem de falta de visibilidade que as limita à análise dos terceiros diretos, em detrimento de uma perspetiva mais alargada das relações existentes na cadeia de fornecimento. Esta limitação reduz a capacidade de captar fenómenos de propagação do ciber-risco ao longo da cadeia de fornecimento. Assim, falta uma solução capaz de contemplar, de forma integrada, tanto o ciber-risco intrínseco, quanto o risco de terceiros e da cadeia de fornecimento, tendo por base apenas informação pública.

Assim sendo, embora já existam soluções capazes de avaliar o ciber-risco de uma organização e o risco de terceiros através do recurso a informação pública, estas não o fazem de forma holística e consolidada, mas sim ou num eixo ou no outro. Por isso, acabam por ter claras limitações de visibilidade relativas aos terceiros de uma organização e um foco extremamente específico nas vulnerabilidades identificadas nos ativos expostos à *Internet*, descurando, por exemplo, a informação passível de ser extraída a partir da presença de credenciais corporativas em fugas de dados. Deste modo, persiste na literatura uma lacuna relevante no desenvolvimento de abordagens que integrem, de forma sistemática, outras fontes de informação pública e múltiplos níveis de interdependência organizacional na avaliação do ciber-risco.

3. OSIRIS-GOV

Este capítulo expõe a proposta de solução *Open-Source Intelligence for Resilient Infrastructure and Security in Government Services* (OSIRIS-GOV), que visa endereçar as principais falhas, lacunas e limitações identificadas no levantamento do estado da arte, ao mapear automaticamente os terceiros diretos e indiretos das organizações para os considerar na avaliação do ciber-risco, unicamente com base em informação pública.

3.1. Enquadramento

De acordo com o levantamento do estado da arte anteriormente efetuado, foram identificadas três principais lacunas nas abordagens estudadas que se propõem a aferir o ciber-risco de uma organização, tendo por base informação pública.

Em primeiro lugar, apesar de os mais recentes estudos e relatórios demonstrarem que os terceiros têm vindo a assumir um papel cada vez mais significativo nos ciberincidentes que afetam as organizações, esta realidade não se encontra devidamente refletida nas soluções analisadas para avaliação do ciber-risco, conforme evidenciado nos trabalhos discutidos anteriormente. De facto, estas soluções ou não contemplam o risco de terceiros, ou, quando o fazem, fazem-no exclusivamente, isto é, têm como único propósito a avaliação do risco associado aos terceiros, fornecedores e prestadores de serviços diretos de cada organização, desconsiderando o ciber-risco da própria organização. Deste modo, existe uma tendência no sentido de dissociar a análise do ciber-risco intrínseco à própria organização da análise do risco de terceiros, pelo que faz falta uma visão integrada de ambas as dimensões.

Em segundo lugar, embora a literatura destaque a crescente importância de, na avaliação do risco de terceiros, serem igualmente considerados os terceiros, fornecedores e prestadores de serviços indiretos, isto é, ao longo da cadeia de fornecimento, dado o seu carácter sistémico e o potencial de propagação do ciber-risco, a generalidade das abordagens propostas não contempla, na prática, isto mesmo. Efetivamente, as abordagens puramente dedicadas ao risco de terceiros tendem a excluir a identificação ou o mapeamento dos terceiros, fornecedores e prestadores de serviços indiretos, ou seja, para além de um primeiro nível de relação direta com a organização. Por isso, ainda que já tenham sido teorizadas algumas tentativas, acabam por não ser aplicados modelos de propagação do ciber-risco ao longo da cadeia de fornecimento.

Em terceiro lugar, na vasta maioria das soluções estudadas, o principal elemento aferidor do ciber-risco incide sobre as vulnerabilidades existentes nos ativos expostos à *Internet*. Esta abordagem, embora amplamente validada na literatura científica recolhida, contrasta com as evidências apresentadas nos relatórios mais recentes, que apontam para a relevância de outros fatores com impacto considerável no ciber-risco organizacional, entre os quais se inclui a presença de credenciais corporativas em fugas de dados. No entanto, a generalidade das abordagens estudadas continua a privilegiar a identificação de vulnerabilidades técnicas, não incorporando, de forma consistente, as fugas de dados na avaliação do ciber-risco. Assim, atendendo ao crescente impacto das credenciais enquanto elemento agravador do ciber-risco organizacional, optou-se por privilegiar as fugas de dados com informação sobre credenciais corporativas expostas como um dos principais eixos de análise, tanto de uma organização, como ao longo da cadeia de fornecimento.

Assim sendo, a solução proposta pretende endereçar estas três lacunas do estado da arte, ao: (1) considerar o risco de terceiros na avaliação do ciber-risco organizacional, (2) mapear automaticamente – usando a informação que está disponível publicamente – a cadeia de fornecimento, identificando os terceiros diretos e indiretos e (3) aferir o ciber-risco com base na exposição em fugas de dados das credenciais corporativas de cada organização.

Esta preferência pela exposição de credenciais em fugas de dados e pelo risco de terceiros enquanto elementos centrais à determinação do ciber-risco organizacional deve-se, essencialmente, ao resultado das falhas, lacunas e limitações identificadas após o levantamento do estado da arte e encontra suporte nas conclusões vertidas no já mencionado DBIR [44]. Conforme aludido anteriormente, este relatório retrata o abuso de credenciais como o principal vetor de acesso inicial nos ciberataques às organizações – com 22% de prevalência nos incidentes considerados – e o envolvimento ou participação de um terceiro em cerca de 30% das ocorrências, pelo que estes dois fatores devem ser considerados na aferição do ciber-risco que impende sobre as organizações.

Numa visão de alto-nível, a Figura 3 ilustra a diferença entre as soluções tradicionais de avaliação do ciber-risco organizacional ou do risco de terceiros, em comparação com a abordagem proposta para o OSIRIS-GOV.

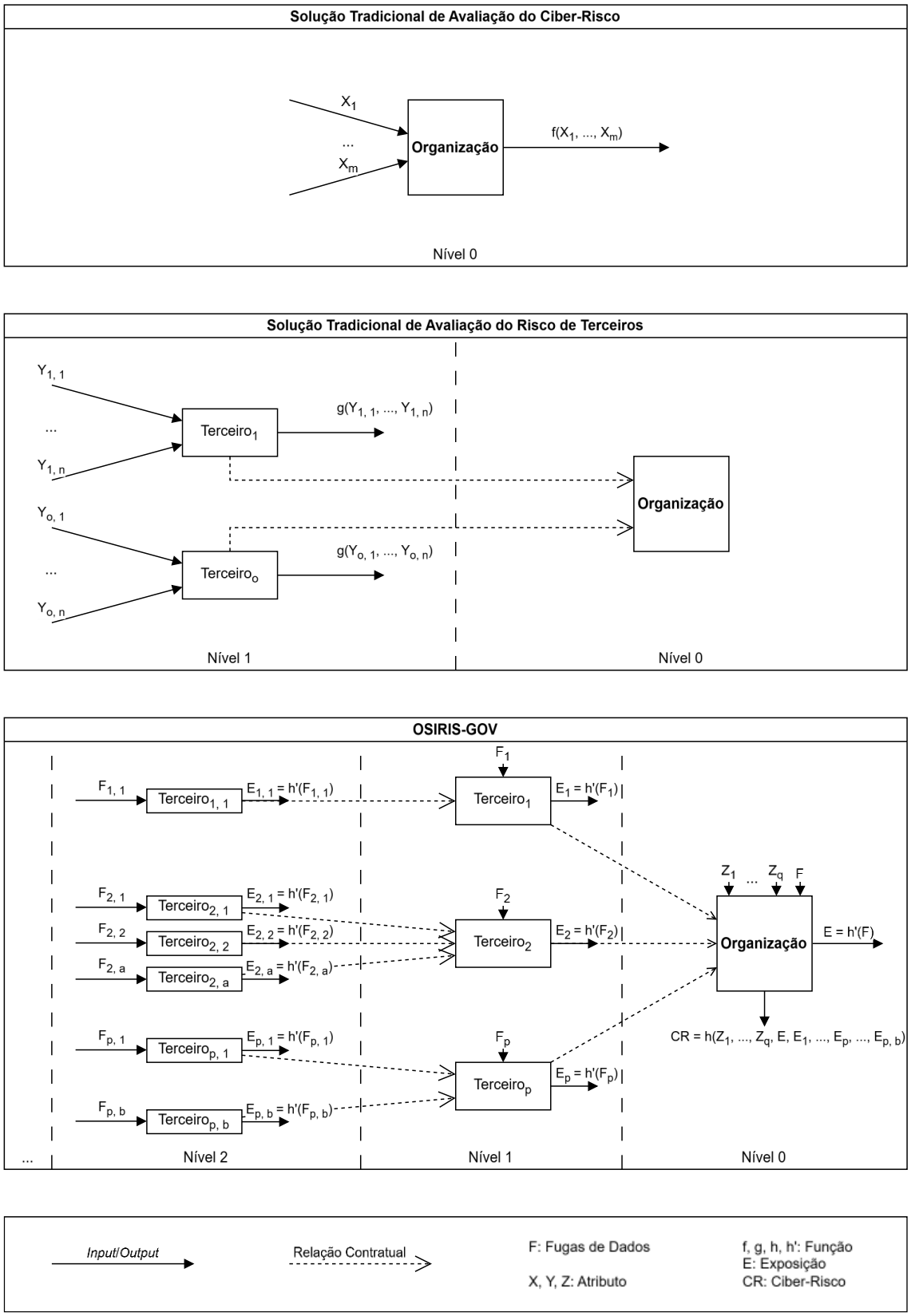

Figura 3 – Comparação entre as Soluções de Avaliação do Ciber-Risco ou do Risco de Terceiros com o OSIRIS-GOV

Tal como a Figura 3 pretende transmitir, o OSIRIS-GOV distingue-se das demais soluções de avaliação do ciber-risco e do risco de terceiros. Conforme ilustrado visualmente na primeira secção da figura, as soluções de avaliação do ciber-risco tendem a considerar apenas um determinado conjunto de atributos da organização – representados por X_1 a X_m – para a avaliação do ciber-risco, isto é, para o cálculo de $f(X_1, \dots, X_m)$. Estes atributos correspondem a determinadas características ou propriedades da organização em análise, que, na maior parte dos casos identificados no levantamento do estado da arte, consistem nas vulnerabilidades detetadas nos ativos expostos à *Internet*. No entanto, estas soluções têm associadas duas principais limitações, já previamente elencadas. Em primeiro lugar, limitam-se a considerar as características da própria organização na avaliação do ciber-risco, sem analisar a componente proveniente dos seus terceiros, fornecedores e prestadores de serviços externos, embora todos estes tenham vindo a ter uma participação cada vez mais significativa nos ciberincidentes atuais. Em segundo lugar, a opção pela análise das vulnerabilidades enquanto fator aferidor do ciber-risco contrasta com as evidências expostas pelos mais recentes relatórios de ciberincidentes, que apontam para o facto de este vetor ter vindo a deixar de ser o mais explorado pelos cibercriminosos para a consumação de ciberataques.

Por sua vez, as soluções de avaliação do risco de terceiros – representadas na segunda secção da figura – já consideram os terceiros na análise do ciber-risco, embora ainda o façam de uma forma limitada e diferente da proposta para o OSIRIS-GOV. Estas soluções recolhem e analisam um conjunto de atributos de cada terceiro da organização (identificados por $Y_{o,1}$ a $Y_{o,n}$, sendo o o índice relativo a cada terceiro), com o objetivo de determinar o ciber-risco associado a esse mesmo terceiro (representado por $g(Y_{o,1}, \dots, Y_{o,n})$, para o terceiro de índice o). Contudo, de acordo com o levantamento do estado da arte anteriormente efetuado, a maior parte destas soluções cinge-se à avaliação do risco dos terceiros com relação contratual direta com a organização, o que acarreta, essencialmente, dois problemas. Por um lado, a identificação dos terceiros tende a estar limitada ao primeiro nível da cadeia de fornecimento, sem considerar o potencial impacto que pode vir a ser aportado pelos fornecedores e prestadores de serviços dos níveis subsequentes da cadeia. Por outro lado, estas soluções focam-se, maioritariamente, na avaliação isolada do risco de cada terceiro, em vez de avaliarem de que forma é que esse risco contribui para o ciber-risco da organização em análise, que depende desses terceiros.

Finalmente, o OSIRIS-GOV – ilustrado na terceira secção da figura – adota uma abordagem distinta, tanto das soluções de avaliação do ciber-risco, quanto das soluções de avaliação do risco de terceiros. A solução proposta vai além do primeiro nível da cadeia de fornecimento para mapear, a par deste, os restantes terceiros que, por via de um efeito de propagação do risco, podem contribuir para o ciber-risco da organização em análise. Para calcular este contributo, o OSIRIS-GOV recolhe informação sobre as fugas de dados associadas a todas as entidades identificadas na cadeia de fornecimento, incluindo a própria organização em análise. Neste contexto, uma fuga de dados corresponde à exposição ou divulgação indevida de informação sensível, sendo que a informação recolhida pelo OSIRIS-GOV consiste na presença de credenciais corporativas – isto é, combinações de um endereço de *e-mail* e uma palavra-passe, cujo domínio do endereço de *e-mail* corresponde ao domínio da organização – nestas fugas de dados. Esta informação (representada por F) é utilizada para determinar o valor da exposição (E) de cada entidade integrante da cadeia de fornecimento, através da aplicação de uma determinada função h'. Deste modo, o OSIRIS-GOV trata a presença de credenciais corporativas em fugas de dados enquanto fator aferidor da exposição para o cálculo do ciber-risco, em detrimento das tradicionais vulnerabilidades nos ativos expostos à *Internet*. Posteriormente, todos os valores de exposição são combinados com os atributos demográficos (Z_1 a Z_q) da organização em análise, para obtenção do respetivo ciber-risco, representado pela aplicação da função h a toda a informação recolhida, ou seja, à exposição da cadeia de fornecimento (incluindo a própria organização) e aos atributos demográficos da organização em análise. Com isto, o OSIRIS-GOV vai além do que é efetuado pelas outras soluções, ao considerar também o risco proveniente dos terceiros diretos e indiretos através da respetiva exposição por credenciais, a par dos atributos relativos à própria organização, numa perspetiva que se pretende mais alinhada com as principais causas dos ciberincidentes atuais.

Note-se que, embora, por limitações gráficas, a representação da Figura 3 termine no segundo nível da cadeia de fornecimento, o OSIRIS-GOV permite percorrer os restantes níveis publicamente observáveis da cadeia para identificar os terceiros, fornecedores e prestadores de serviços da organização a considerar, bem como a exposição associada e o seu contributo na avaliação do ciber-risco organizacional.

Em síntese, o OSIRIS-GOV estrutura-se em torno de três eixos fundamentais: o mapeamento automático da cadeia de fornecimento publicamente observável, a integração do risco de terceiros na avaliação do ciber-risco organizacional e a análise da exposição de credenciais em fugas de dados enquanto fator aferidor do ciber-risco.

Pelas razões inicialmente explicitadas, referentes à emergência e à importância da cibersegurança no setor da administração pública, a conceção da solução focou-se especificamente neste setor e, em particular, na administração pública local, o que justifica a designação OSIRIS-GOV.

3.2. Requisitos

De acordo com este enquadramento, foram definidos os principais requisitos aplicáveis ao OSIRIS-GOV, com o intuito de ir ao encontro do objetivo proposto para a solução.

Essencialmente, o que se pretende é que a solução proposta seja capaz de suprir as principais falhas, lacunas e limitações identificadas no levantamento do estado da arte quanto às soluções de avaliação do ciber-risco organizacional, com base em informação pública. Nesta medida, o OSIRIS-GOV deve ser uma solução que, através do recurso exclusivo a informação disponível publicamente, avalie o ciber-risco de uma determinada organização, introduzida pelo utilizador. Nesta avaliação, deve ser considerado o contributo dos terceiros, fornecedores e prestadores de serviços da organização em análise, quer tenham uma relação contratual direta com a mesma, quer sejam terceiros indiretos, isto é, pertencentes à cadeia de fornecimento, mas num nível mais distante. De maneira a refletir adequadamente os diferentes serviços prestados pelos terceiros e o seu potencial impacto na propagação do risco, a solução deve ponderar de forma diferente o impacto das entidades integrantes da cadeia de fornecimento, conforme o tipo de serviços prestados. Além disto, com o intuito de utilizar, para a avaliação do ciber-risco, os principais fatores que têm vindo a ser explorados para a concretização dos mais recentes ciberincidentes, a solução deve contemplar na análise a exposição de credenciais corporativas em fugas de dados. Finalmente, na tentativa de obter uma avaliação do ciber-risco que seja, em simultâneo, independente da opinião de especialistas, objetiva e facilmente entendível, o resultado retornado deve consistir na probabilidade de ocorrência de um ciberincidente que afete a organização, enquanto descritor objetivo do ciber-risco avaliado.

Para isto, elencaram-se os oito requisitos funcionais (RF) que se apresentam na Tabela 13. Estes requisitos definem as principais funcionalidades e capacidades que o OSIRIS-GOV deve possuir, de modo a satisfazer as necessidades sintetizadas anteriormente.

Requisito	Descrição
RF1	A solução deve permitir a introdução de uma organização para avaliação do ciber-risco, a partir de um identificador único.
RF2	A solução deve processar informação relativa a atributos demográficos da organização para avaliação do ciber-risco.
RF3	A solução deve processar informação relativa à exposição em fugas de dados com credenciais da organização para avaliação do ciber-risco.
RF4	A solução deve permitir englobar as entidades pertencentes à cadeia de fornecimento da organização na avaliação do ciber-risco.
RF5	A solução deve ponderar de forma diferente as entidades pertencentes à cadeia de fornecimento, conforme o potencial impacto que um eventual comprometimento poderá ter no ciber-risco da organização.
RF6	A solução deve processar informação relativa à exposição em fugas de dados com credenciais das entidades pertencentes à cadeia de fornecimento da organização para avaliação do ciber-risco.
RF7	A solução deve avaliar o ciber-risco da organização introduzida, considerando os atributos da organização e a exposição em fugas de dados da própria organização e da cadeia de fornecimento.
RF8	A solução deve retornar como resultado uma pontuação correspondente ao ciber-risco avaliado para a organização introduzida, descrito pela probabilidade de ocorrência de um ciberincidente.

Tabela 13 – Requisitos Funcionais

Paralelamente, o funcionamento da solução requer o cumprimento de determinadas condições. De acordo com o enquadramento delineado, a avaliação do ciber-risco deve basear-se apenas em informação pública. Além disto, o OSIRIS-GOV deve oferecer garantias de desempenho, disponibilidade, privacidade e usabilidade, por forma a garantir a sua aplicação prática, assim como assegurar a sua manutenibilidade e extensibilidade, para ser facilmente adaptável e ter capacidades de evolução futura.

Assim, a Tabela 14 lista os sete requisitos não funcionais (RNF) da solução, que estabelecem as condições e restrições que o OSIRIS-GOV deve satisfazer.

Requisito	Descrição
RNF1	A solução deve recolher e tratar exclusivamente informação pública.
RNF2 (Desempenho)	A solução deve retornar o resultado para o ciber-risco avaliado para uma organização em menos de cinco minutos.
RNF3 (Disponibilidade)	A solução deve estar permanentemente disponível para ser executada localmente, a pedido do utilizador.
RNF4 (Privacidade)	A solução deve evitar a recolha de dados sensíveis, ou minimizá-la ao estritamente necessário.
RNF5 (Usabilidade)	A solução deve ser facilmente utilizável por parte dos utilizadores.
RNF6 (Manutenibilidade)	A solução deve ser devidamente documentada, de maneira a suportar a sua manutenção e eventual evolução futura.
RNF7 (Extensibilidade)	A solução deve ser desenvolvida numa arquitetura modular, facilmente extensível a diferentes fontes de informação.

Tabela 14 – Requisitos Não Funcionais

3.3. Arquitetura

No contexto do enquadramento anteriormente descrito e para cumprir os requisitos propostos, definiu-se a arquitetura da solução OSIRIS-GOV. Assim, a Figura 4 apresenta uma visão de alto-nível da arquitetura delineada.

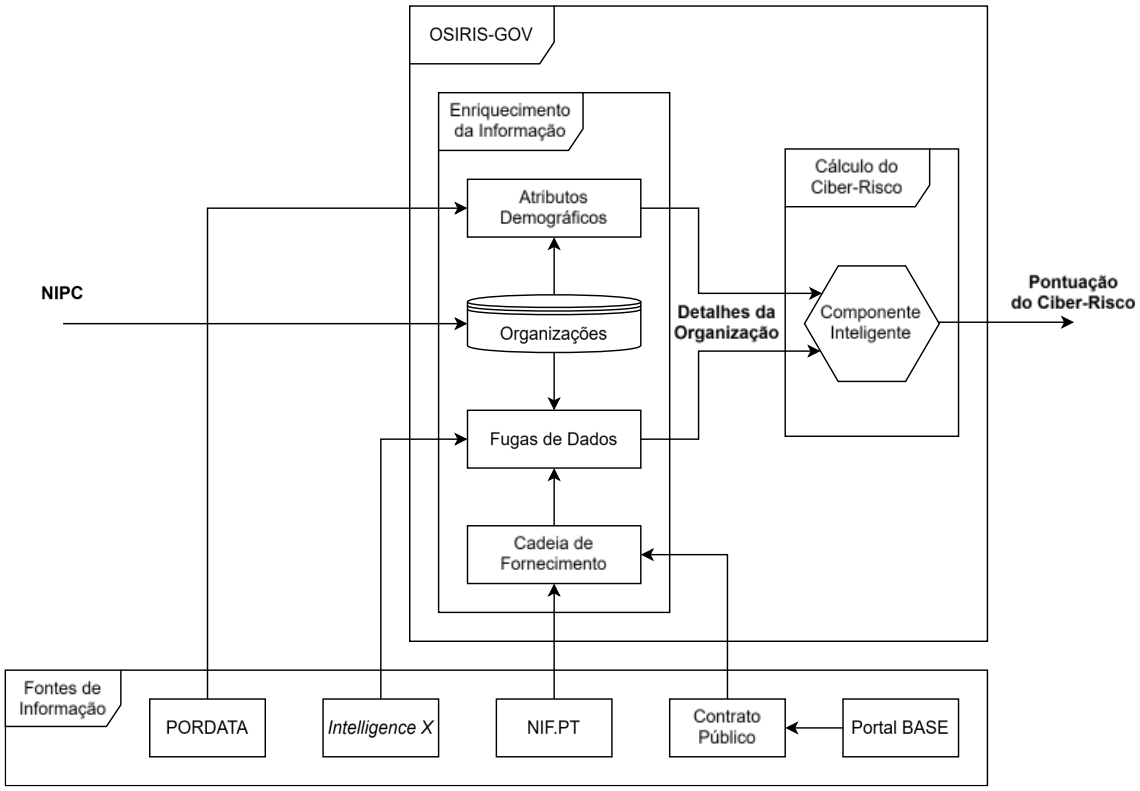

Figura 4 – Diagrama de Arquitetura

Conforme ilustra a Figura 4, o OSIRIS-GOV é constituído por dois módulos internos (“Enriquecimento da Informação” e “Cálculo do Ciber-Risco”), suportados por um conjunto de fontes de informação externas, seguindo uma arquitetura modular orientada pelos fluxos de dados.

O valor de entrada na solução é o Número de Identificação de Pessoa Coletiva (NIPC) da organização em análise. A partir da introdução deste valor, é identificada a organização em causa e são recolhidos os seus dados demográficos, assim como informação sobre fugas de dados com credenciais corporativas da própria organização e das entidades integrantes da sua cadeia de fornecimento, mapeada automaticamente pelo próprio OSIRIS-GOV. Estes atributos constituem os “Detalhes da Organização”, que são transmitidos internamente entre ambos os módulos da solução. Após o processamento desta informação, a solução retorna como valor de saída a probabilidade de ocorrência de um ciberincidente na organização em análise num dado intervalo de tempo futuro, sendo esta a pontuação descritora do ciber-risco calculado pelo OSIRIS-GOV para a organização.

Na parte inferior do diagrama, encontram-se as fontes de informação externas que disponibilizam de forma publicamente acessível os dados recolhidos pela solução. Estas fontes são a PORDATA[49], o *Intelligence X*, o NIF.PT[50], o Contrato Público[51] e o Portal BASE[52], tendo sido escolhidas pela adequabilidade ao contexto da administração pública local, mas sendo facilmente substituídas ou complementadas por outras, dependendo do contexto pretendido. A primeira permite a recolha dos atributos demográficos da organização e a segunda disponibiliza informação sobre fugas de dados com credenciais expostas, enquanto as restantes são consultadas para efetuar o mapeamento da cadeia de fornecimento da organização em análise. De acordo com os requisitos previamente definidos – que estabelecem que o cálculo do ciber-risco organizacional se deve basear nos atributos demográficos da organização, juntamente com a exposição em fugas de dados das suas credenciais e das credenciais da cadeia de fornecimento em que se insere –, estes são os dados necessários para o funcionamento do OSIRIS-GOV.

No interior do OSIRIS-GOV, expõem-se os dois módulos da solução: o primeiro (do lado esquerdo) é responsável pelo cruzamento e enriquecimento da informação previamente recolhida e o segundo (do lado direito) permite o cálculo do ciber-risco organizacional. Sucintamente, o módulo “Enriquecimento da Informação” interage com as fontes de informação externas para recolher os dados anteriormente mencionados, cruzando e organizando a informação recebida para a transmitir adequadamente ao módulo seguinte. Esta informação constitui os “Detalhes da Organização”. A partir destes dados, o módulo “Cálculo do Ciber-Risco” efetua o seu processamento interno através de um “Componente Inteligente” e devolve a probabilidade de que um ciberincidente venha a ocorrer na organização em causa, enquanto valor que descreve, de forma concreta e objetiva, o resultado da avaliação do ciber-risco efetuada pela solução.

As fontes de informação e os dois módulos internos do OSIRIS-GOV são descritos, por esta ordem, nas três subsecções seguintes.

[49] https://www.pordata.pt/
[50] https://www.nif.pt/
[51] https://contratopublico.pt/
[52] https://www.base.gov.pt/

3.3.1. Fontes de Informação

Conforme introduzido anteriormente, o OSIRIS-GOV é suportado por cinco fontes de informação externas: a PORDATA, o *Intelligence X*, o NIF.PT, o Contrato Público e o Portal BASE.

A PORDATA – Base de Dados de Portugal Contemporâneo, desenvolvida pela Fundação Francisco Manuel dos Santos – recolhe, organiza, sistematiza e divulga informação sobre diversas áreas da sociedade europeia, com especial foco em Portugal e nos municípios portugueses [65]. Esta fonte de informação pública agrega estatísticas provenientes de múltiplas fontes oficiais e certificadas, disponibilizando-as de forma clara, acessível e gratuita a todos os utilizadores [65]. Sendo o âmbito do OSIRIS-GOV particularmente focado no contexto da administração pública local portuguesa, a PORDATA foi a opção escolhida para a recolha dos atributos demográficos relativos a cada organização em análise.

O *Intelligence X* – identificado no levantamento do estado da arte – é a fonte que disponibiliza informação sobre fugas de dados com presença de credenciais corporativas expostas. Esta fonte funciona como um motor de pesquisa que arquiva e indexa dados para os disponibilizar publicamente, através de uma API, sendo estes dados organizados e agrupados em *buckets*, de acordo com a sua proveniência. Para o efeito, foram selecionados para consulta três *buckets* distintos: (1) o *bucket leaks.logs*, que contém fugas de dados provenientes de dispositivos infetados por *infostealers*, um tipo de *malware* (*software* malicioso) desenvolvido com o propósito específico de recolher dados sensíveis das máquinas infetadas, como as credenciais nelas armazenadas, (2) o *bucket leaks.private.comb*, que alberga resultados de compilações de listas com elevados volumes de credenciais expostas em fugas de dados e (3) o *bucket leaks.private.general*, ao qual ficam associados outros tipos de fugas de dados, também com credenciais expostas, mas que não se enquadrem em nenhum dos outros dois *buckets* [66].

O NIF.PT é um serviço gratuito que permite a pesquisa do Número de Identificação Fiscal (NIF) português, possibilitando a procura, descoberta e consulta do NIPC – enquanto identificador fiscal de pessoas coletivas – de empresas, através da sua API [67]. Por isso, foi a fonte de informação consultada de forma a efetuar o mapeamento entre o NIPC de cada organização – para pesquisa no Contrato Público – e o respetivo domínio – para pesquisa no *Intelligence X*.

O Portal BASE é o portal português destinado a "divulgar informação pública sobre os contratos públicos sujeitos ao regime do Código dos Contratos Públicos" [68]. Nesta medida, tem como principal função "centralizar a informação sobre os contratos públicos celebrados em Portugal, configurando um espaço virtual no qual são publicitados os elementos referentes à formação e execução dos contratos públicos, permitindo o seu acompanhamento e monitorização" [68]. Assim sendo, esta fonte de informação perfilou-se como uma solução natural para recolher informação capaz de contribuir para a identificação automática dos terceiros, fornecedores e prestadores de serviços das organizações da administração pública, com o intuito de mapear a respetiva cadeia de fornecimento, tendo por base as relações contratuais publicamente divulgadas.

Para tal, foi considerada a utilização da API[53] disponibilizada desde maio de 2025 pelo Portal BASE para consulta e extração de grandes volumes de dados. No entanto, o acesso a esta API é condicionado ao registo e à prévia autorização por parte do Instituto dos Mercados Públicos, do Imobiliário e da Construção (IMPIC), que não emitiu resposta em tempo útil à solicitação efetuada neste sentido. Como tal, acrescendo a isto o facto de que a execução de pesquisas no Portal BASE apresenta um desempenho consideravelmente fraco em termos de tempo de execução/resposta – frequentemente superior a cinco segundos e, esporadicamente, superior a 50 segundos –, foi procurada uma solução alternativa para a recolha dos dados deste mesmo portal.

Nesta medida, foram comparadas três soluções capazes de recolher dados do Portal BASE: *base-gov-pt-scraper*[54], *base.gov.pt*[55] e o Contrato Público[56], todas disponíveis publicamente através dos respetivos repositórios. Dos testes efetuados, a terceira solução foi aquela que apresentou um melhor desempenho, pelo que se optou pela sua utilização.

O Contrato Público "é um serviço de pesquisa de contratos públicos realizados em Portugal que agrega dados do Portal BASE, disponibilizando-os numa plataforma muito mais usável, rápida e intuitiva", com tempos de resposta na ordem de grandeza dos milissegundos [69]. Para o efeito, o Contrato Público recolhe continuamente dados do Portal BASE, sendo, por isso, adequado para obter a informação necessária ao mapeamento da cadeia de fornecimento, indo ao encontro dos requisitos delineados.

[53] https://www.base.gov.pt/Base4/pt/noticias/2025/api-para-consulta-de-dados-do-portal-base/
[54] https://github.com/jpsgamboa/base-gov-pt-scraper
[55] https://github.com/ajcerejeira/base.gov.pt
[56] https://github.com/chicoferreira/contratopublico

Assim, selecionaram-se estas fontes de informação pública para consulta e recolha dos dados necessários ao cumprimento dos requisitos propostos para o OSIRIS-GOV.

3.3.2. Enriquecimento da Informação

No módulo "Enriquecimento da informação", coexistem quatro componentes: uma base de dados e três integrações com as fontes de informação anteriormente descritas.

A base de dados "Organizações" contém o NIPC, o nome e o domínio das organizações a analisar. Por exemplo, no caso de estas organizações serem as câmaras municipais de Portugal, embora a informação das mesmas também pudesse ser obtida por consulta da fonte de informação NIF.PT, foi recolhida previamente da Associação Nacional de Municípios Portugueses (ANMP), com o intuito de minimizar os pedidos efetuados à API do NIF.PT, por serem extremamente limitados. Esta base de dados é consultada imediatamente após a introdução do valor de entrada (o NIPC) no sistema, de modo a obter o nome e o domínio da entidade correspondente, necessários para efetuar a pesquisa nas fontes de informação.

O nome da organização é utilizado para consulta dos atributos demográficos da mesma, pelo componente "Atributos Demográficos". O recurso à informação demográfica de cada organização para avaliar o respetivo ciber-risco deve-se às propostas de alguns autores identificados na literatura, como Sarabi *et al.* [42], que utilizam os atributos demográficos presentes na VCDB para este mesmo efeito. Em particular, são utilizados os seguintes atributos da VCDB, referentes a cada organização: o país, o número de colaboradores, a indústria ou o setor de atividade e a indicação de pertença (ou não) ao setor da administração pública. Embora o sistema desenvolvido seja extensível e generalizável para ser aplicado a qualquer organização, no caso particular da introdução de uma câmara municipal, é apenas recolhida a informação – proveniente da PORDATA – relativa ao número de colaboradores da organização, tendo os restantes atributos (país, indústria/setor e indicação de pertença ao setor da administração pública) valores fixos, definidos, respetivamente, como "Portugal", "Administração Pública" e "Verdadeiro". Estes atributos fazem parte da informação a processar, posteriormente, pelo "Componente Inteligente", para cálculo do ciber-risco organizacional.

Por sua vez, o domínio da organização é utilizado para recolha de informação sobre as fugas de dados com credenciais expostas associadas a esse domínio, pelo componente denominado "Fugas de Dados". Para isso, são procuradas no *Intelligence X*, num intervalo de tempo parametrizável, todas as ocorrências de fugas de dados que contenham endereços de *e-mail* associados ao domínio da organização, juntamente com uma palavra-passe. Esta informação é mapeada diretamente nos atributos *leaks.logs*, *leaks.private.comb* e *leaks.private.general*, cujos nomes correspondem aos três *buckets* do *Intelligence X* a partir dos quais é recolhida a informação. Assim, cada um destes três atributos contém o valor absoluto de ocorrências de fugas de dados atribuídas pelo *Intelligence X* a cada *bucket* correspondente e com credenciais expostas relativas à organização em causa, no intervalo de tempo considerado. Esta informação permite aferir a exposição por credenciais da organização em análise, necessária para o cálculo da pontuação descritora do respetivo ciber-risco.

Finalmente, o componente "Cadeia de Fornecimento" é responsável por pesquisar – no Contrato Público – as entidades que fazem parte da cadeia de fornecimento da organização em análise, isto é, os terceiros, fornecedores e prestadores de serviços com relação direta ou indireta com a organização, identificada pelo seu NIPC. Com base na informação recolhida, é modelado um grafo correspondente à cadeia de fornecimento da organização, sendo obtido o domínio de cada entidade integrante da cadeia a partir do NIF.PT. As entidades pertencentes a este grafo são, posteriormente, pesquisadas pelo componente "Fugas de Dados" quanto à existência de fugas de dados com credenciais associadas, somando os valores resultantes nos mesmos três atributos anteriores (*leaks.logs*, *leaks.private.comb* e *leaks.private.general*), com uma ponderação configurável conforme a respetiva relação contratual e o seu significado. Desta forma, estes três atributos passam a contemplar a exposição por credenciais de toda a cadeia de fornecimento da organização.

Para isso, a cadeia de fornecimento é modelada como um grafo dirigido pesado $G = (V, E)$, no qual o conjunto de vértices V corresponde às entidades que fazem parte da cadeia de fornecimento – incluindo a organização em análise e os seus terceiros, fornecedores e prestadores de serviços, diretos ou indiretos – e o conjunto de arestas E representa as relações contratuais estabelecidas entre estas entidades. Neste grafo, cada aresta $e_{ij} \in E$ traduz a existência de uma relação contratual entre duas entidades $i \in V$ e $j \in V$, sendo i a entidade adjudicatária/contratada e j a respetiva entidade adjudicante/contratante.

De forma a traduzir o sentido da propagação do risco, o sentido de cada aresta é definido da entidade adjudicatária/contratada para a entidade adjudicante/contratante, com o intuito de representar o potencial impacto que a primeira pode exercer sobre a segunda no contexto da propagação do ciber-risco ao longo da cadeia de fornecimento. O peso w_e de cada aresta $e \in E$ é um valor entre zero e um que procura traduzir o potencial impacto para a entidade de destino em caso de comprometimento da entidade de origem.

O objetivo com a elaboração deste grafo passa por permitir avaliar o potencial contributo de cada terceiro, fornecedor ou prestador de serviços – enquanto parte integrante da cadeia de fornecimento – no ciber-risco da organização em análise. Nesta medida, depois de obter, para cada entidade/nó do grafo, os valores correspondentes aos atributos *leaks.logs*, *leaks.private.comb* e *leaks.private.general* (com a frequência absoluta de fugas de dados com credenciais expostas associadas ao domínio da respetiva entidade), estes mesmos valores são propagados ao longo da cadeia de fornecimento. Esta propagação visa estimar o impacto que cada terceiro – direto ou indireto – pode ter no ciber-risco da organização em análise, refletindo não só o tipo de serviços prestados, mas também a proximidade desse terceiro em relação à organização em avaliação.

Para isso, considera-se, no grafo que modela a cadeia de fornecimento, que um caminho l_{ij} entre duas entidades $i \in V$ e $j \in V$ é um percurso desde o vértice i até ao vértice j que, respeitando o sentido das arestas, só pode passar uma vez por cada vértice do grafo. O peso total de um caminho l_{ij} – representado por $W(l_{ij})$ – é dado pelo produto dos pesos de todas as arestas constituintes do caminho. Com base nesta definição, é possível propagar o valor dos atributos *leaks.logs*, *leaks.private.comb* e *leaks.private.general* ao longo da cadeia de fornecimento, desde cada vértice do grafo até à organização em análise. Sendo $x \in V$ a organização em análise e l_{vx} o caminho com maior peso desde cada entidade $v \in V$ até à organização x, o valor final de cada atributo do triplo $L = (leaks.logs, leaks.private.comb, leaks.private.general)$ é dado pela expressão da Equação (4), na qual L_x e L_v correspondem, respetivamente, aos valores dos atributos deste mesmo triplo L para a organização em análise x e para cada entidade v.

$$L = L_x + \sum_{v \in V \setminus \{x\}} L_v \times W(l_{vx}) \tag{4}$$

Desta forma, é acumulada a soma dos valores destes três atributos de cada entidade nos mesmos três atributos para a organização em análise, de forma ponderada conforme os pesos dos caminhos correspondentes. Com isto, garante-se que, na medida em que for possível identificar a cadeia de fornecimento da organização a avaliar, as entidades da cadeia são consideradas na avaliação do ciber-risco, refletindo sempre as diferentes relações contratuais existentes e o potencial impacto associado.

Após o processamento dos quatro componentes deste módulo, a informação resultante é passada ao módulo de cálculo do ciber-risco, sob a forma de um conjunto de atributos, designado por "Detalhes da Organização". Estes atributos incluem não só informação demográfica da organização (país, número de colaboradores, indústria/setor e indicação de pertença ao setor da administração pública), mas também informação sobre as fugas de dados com credenciais corporativas expostas, tanto da organização, como da respetiva cadeia de fornecimento (*leaks.logs*, *leaks.private.comb* e *leaks.private.general*), todos requeridos para a avaliação do ciber-risco organizacional.

Em resumo, o módulo de enriquecimento da informação é responsável por recolher e tratar toda a informação considerada necessária para o cálculo do ciber-risco, transmitindo-a assim que o processamento esteja concluído.

3.3.3. Cálculo do Ciber-Risco

O módulo de cálculo do ciber-risco é constituído por um componente inteligente capaz de ingerir toda a informação que lhe é transmitida, devolvendo, como resultado, uma pontuação para o ciber-risco avaliado para a organização em causa, traduzido através da probabilidade estimada de ocorrência de um ciberincidente.

A única função deste módulo passa por analisar os "Detalhes da Organização" (enviados pelo módulo anterior) e retornar o resultado pretendido. Para este efeito, segue-se uma abordagem de aprendizagem automática supervisionada, devido à necessidade de identificar padrões e relações entre múltiplos atributos organizacionais – tanto demográficos, quanto relativos à exposição por credenciais –, cuja influência conjunta na ocorrência de ciberincidentes poderia ser difícil de modelar através de abordagens heurísticas ou determinísticas. Desta forma, o recurso a um classificador inteligente está alinhado com as soluções analisadas na revisão de literatura, que procuram tirar proveito das capacidades de aprendizagem automática e inferência sobre grandes volumes de informação, tendo em vista a avaliação do ciber-risco organizacional.

O processo de cálculo do ciber-risco inicia-se com a receção do conjunto de atributos que caracterizam a organização, incluindo a informação demográfica e os indicadores da exposição por credenciais da própria organização e da respetiva cadeia de fornecimento. Estes atributos são utilizados como variáveis de entrada para o classificador inteligente, que analisa os padrões aprendidos durante a fase de treino a partir de exemplos históricos de organizações com diferentes características demográficas, níveis de exposição por credenciais e ocorrência de ciberincidentes. Com base nestas relações previamente aprendidas, o classificador estima a probabilidade de a organização em causa vir a sofrer um ciberincidente durante o horizonte temporal considerado, produzindo a pontuação de ciber-risco que constitui o resultado deste módulo.

Assim, o componente inteligente processa a informação que lhe é passada e, tal como nas abordagens – identificadas na literatura recolhida – de Liu *et al.* [37] e de Sarabi *et al.* [42], retorna uma estimativa da probabilidade de ocorrência de um ciberincidente num intervalo de tempo futuro, sendo este o descritor do ciber-risco avaliado. Em particular, pretende-se que o componente consiga inferir de que forma é que diferentes combinações dos atributos recolhidos podem estar relacionadas com a ocorrência de ciberincidentes, permitindo transformar um conjunto de indicadores heterogéneos numa estimativa do ciber-risco. Deste modo, o resultado produzido pelo módulo de cálculo do ciber-risco e, consequentemente, pelo OSIRIS-GOV, é concreto e diretamente interpretável, sem depender de escalas subjetivas ou de níveis de risco definidos por especialistas.

Assim sendo, este módulo materializa o objetivo central do OSIRIS-GOV, ao converter os diversos indicadores recolhidos ao longo do processo numa estimativa clara e objetiva do ciber-risco organizacional.

4. Implementação

Este capítulo detalha a implementação de cada módulo e dos respetivos componentes da solução OSIRIS-GOV proposta anteriormente, com particular foco na descrição do componente inteligente, desde a construção do conjunto de dados até à obtenção de resultados para validação, passando pelas fases intermédias de treino e teste do mesmo.

4.1. Visão Geral do Sistema

A implementação do sistema OSIRIS-GOV seguiu a arquitetura modular delineada anteriormente.

De acordo com a arquitetura estabelecida, o OSIRIS-GOV é constituído por dois módulos internos: o módulo de enriquecimento da informação e o módulo de cálculo do ciber-risco. Sucintamente, o primeiro é responsável por comunicar com as fontes de informação pública de maneira a recolher todos os dados necessários à avaliação do ciber-risco organizacional por parte do segundo.

Para a implementação de ambos os módulos, foi utilizada a linguagem de programação *Python*[57]. Esta foi a opção escolhida dada a maior familiaridade com a mesma e a simplicidade de operação, facilitando a concretização da arquitetura proposta para o OSIRIS-GOV.

A interação com o utilizador, tanto para introdução do valor de entrada no sistema, como para retorno do valor de saída do OSIRIS-GOV, é realizada através do terminal. A comunicação entre ambos os módulos, isto é, a passagem de informação, após o respetivo enriquecimento, para o cálculo do ciber-risco, é efetuada através da invocação de funções. Assim, depois da introdução do valor de entrada por parte do utilizador, o programa *Python* encarrega-se de chamar as funções pertencentes ao módulo de enriquecimento da informação para recolher os dados a partir das fontes de informação externas. Concluído o processamento deste módulo, a informação é transmitida ao módulo de cálculo do ciber-risco, que, por sua vez, retorna o valor a apresentar ao utilizador, com a pontuação correspondente ao ciber-risco avaliado, que se descreve na probabilidade de ocorrência de um ciberincidente.

[57] https://www.python.org/

O diagrama de sequência da Figura 5 apresenta a visão geral deste fluxo de informação.

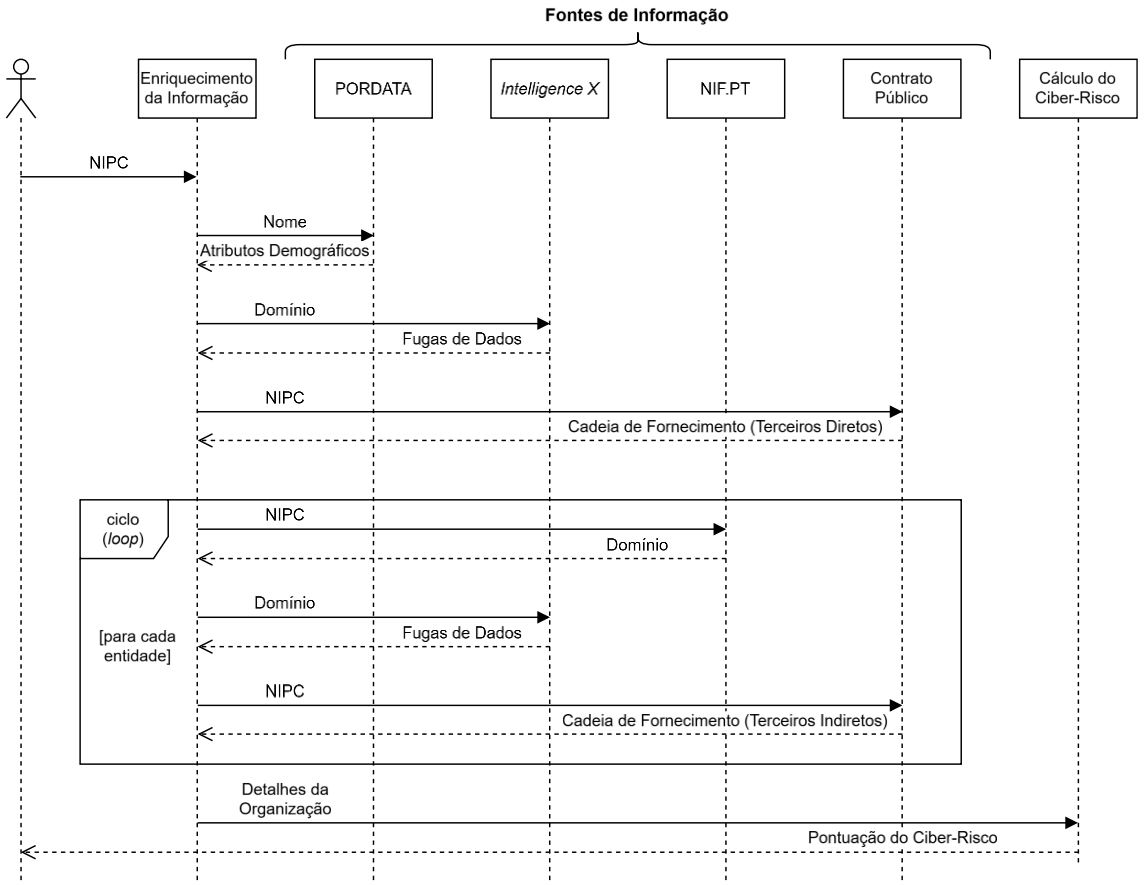

Figura 5 – Diagrama de Sequência

No topo da Figura 5, encontram-se representadas sete entidades: o utilizador, as quatro fontes de informação consultadas pelo OSIRIS-GOV – a PORDATA, o *Intelligence X*, o NIF.PT e o Contrato Público – e os dois módulos internos do sistema, isto é, o módulo de enriquecimento da informação e o módulo de cálculo do ciber-risco. Embora tenha sido mencionado no diagrama de arquitetura da Figura 4, o Portal BASE não se encontra representado neste diagrama de sequência, visto que não interage diretamente com nenhum dos componentes internos do OSIRIS-GOV, servindo apenas para a recolha dos dados por parte do Contrato Público.

Tal como mostra o diagrama, o funcionamento do programa inicia-se com a introdução do NIPC da organização em análise por parte do utilizador, sendo esta a única informação fornecida manualmente ao OSIRIS-GOV.

O valor introduzido é processado pelo módulo “Enriquecimento da Informação” que, internamente, o mapeia no nome e no domínio da organização correspondente. Em seguida, estes dados são cruzados com informação proveniente da PORDATA, do *Intelligence X* e do Contrato Público, respetivamente, para obtenção dos atributos demográficos da organização, das fugas de dados com credenciais expostas associadas e das entidades com relação contratual direta, isto é, que fazem parte do primeiro nível da cadeia de fornecimento, sendo estes os terceiros diretos da organização.

Para cada um destes terceiros diretos, é seguido o ciclo representado na figura. Este ciclo consiste, em primeiro lugar, na identificação do domínio de cada terceiro, obtido por comunicação com o NIF.PT. A partir deste domínio, são recolhidas do *Intelligence X* as fugas de dados com credenciais corporativas expostas. Finalmente, a última etapa do ciclo passa por consultar o Contrato Público para obter os terceiros, fornecedores e prestadores de serviços de cada entidade identificada, de maneira a mapear a restante cadeia de fornecimento. Este mesmo ciclo é seguido para as novas entidades identificadas nesta última etapa. Assim, com este processamento iterativo, são obtidos os terceiros indiretos da organização em análise e as fugas de dados com as respetivas credenciais corporativas.

Concluído o ciclo, o módulo “Enriquecimento da Informação” trata internamente a informação recebida, de modo a criar o grafo que mapeia a cadeia de fornecimento e a obter os valores dos atributos já referidos: os demográficos da organização (país, número de colaboradores, indústria/setor e indicação de pertença ao setor da administração pública) e os relativos à exposição por credenciais de toda a cadeia de fornecimento (*leaks.logs*, *leaks.private.comb* e *leaks.private.general*). O conjunto formado por estes dados constitui os “Detalhes da Organização”, que é enviado ao módulo “Cálculo do Ciber-Risco”. Este módulo processa os dados recebidos e retorna ao utilizador a pontuação obtida para o ciber-risco avaliado, ou seja, a probabilidade estimada de ocorrência de um ciberincidente na organização, num intervalo de tempo futuro.

As secções seguintes detalham a implementação de cada um dos módulos internos do OSIRIS-GOV.

4.2. Enriquecimento da Informação

O módulo de enriquecimento da informação é o responsável por comunicar com todas as fontes de informação externas ao OSIRIS-GOV, fazendo-o de forma adaptada a cada uma.

A base de dados "Organizações" encontra-se num ficheiro CSV armazenado localmente, no qual cada linha corresponde a uma organização. Por exemplo, caso as organizações em análise sejam as câmaras municipais portuguesas, este ficheiro contém o NIPC, o nome e o domínio de cada uma, sendo imediatamente consultado após a introdução do valor de entrada pelo utilizador.

Por sua vez, a informação recolhida da PORDATA é armazenada localmente num outro ficheiro CSV, previamente descarregado a partir desta fonte de informação. No mesmo caso das câmaras municipais, esta informação é consultada de modo a obter o número de colaboradores de cada uma a partir do seu nome, sendo este um dos atributos demográficos fornecidos ao módulo de cálculo do ciber-risco organizacional.

A comunicação com o *Intelligence X* – enquanto fonte para consulta de informação sobre fugas de dados com credenciais corporativas expostas – é efetuada através do *Intelligence X Software Development Kit*[58] (SDK). Internamente, este SDK realiza pedidos HTTP à API do *Intelligence X*, divididos em duas fases, sendo que ambos os pedidos requerem autenticação com uma conta gratuita de utilizador do *Intelligence X*. Em primeiro lugar, é enviado um pedido HTTP POST ao *endpoint* de pesquisa (*/intelligent/search*), para iniciar uma pesquisa das ocorrências de fugas de dados nas quais estejam presentes credenciais cujo endereço de *e-mail* está associado a um determinado domínio, enviado no corpo do pedido. É neste pedido que são selecionados os três *buckets* a consultar (*leaks.logs*, *leaks.private.comb* e *leaks.private.general*), bem como o intervalo de tempo para pesquisa dos resultados, conforme a data de indexação pelo *Intelligence X*. A resposta ao pedido é um identificador único da pesquisa efetuada. Em segundo lugar, este identificador é enviado num pedido HTTP GET ao *endpoint* de consulta de estatísticas (*/intelligent/search/statistic*), que retorna, em formato JSON, a contagem (em valor absoluto) de todas as fugas de dados encontradas, discriminadas pelo *bucket* em que se enquadram. Cada um destes valores é tratado como um atributo a fornecer ao módulo de cálculo do ciber-risco.

[58] https://github.com/IntelligenceX/SDK

A interação com a API do NIF.PT é realizada por um pedido HTTP GET por cada organização, no qual é enviado o respetivo NIPC, bem como uma chave de API. A resposta – em formato JSON – consiste numa série de atributos da organização, entre os quais se inclui o endereço de *e-mail* corporativo da mesma, a partir do qual é extraído o domínio associado. Este domínio – necessário para a pesquisa de fugas de dados com presença de credenciais corporativas expostas – é armazenado localmente, de maneira a minimizar futuros pedidos à API.

Por fim, o Contrato Público é consultado através de pedidos HTTP GET à respetiva API. Nestes pedidos, não é necessária autenticação, sendo apenas enviado o NIPC da entidade contratante e o intervalo de tempo para pesquisa dos contratos públicos, conforme a respetiva data de assinatura e formalização. A resposta da API consiste numa lista, em formato JSON, com os diversos contratos que cumprem todos os filtros enviados no pedido. Nestes contratos, consta, entre outras informações, o NIPC de cada entidade contratada, que, posteriormente, é traduzido para um domínio, para serem pesquisadas as fugas de dados com presença de credenciais desse mesmo domínio. Com a informação das relações contratuais, é modelada a cadeia de fornecimento da organização em análise, de acordo com a arquitetura explicitada. Na tentativa de refletir apenas a cadeia de fornecimento atual da organização em análise, isto é, resultante das relações contratuais atualmente em vigor, admitem-se apenas as arestas correspondentes a relações contratuais estabelecidas no período do último ano.

Para a determinação do peso de cada aresta, é considerada a natureza dos serviços prestados pela entidade contratada à entidade contratante, de acordo com o Vocabulário Comum para os Contratos Públicos (CPV). Este sistema – desenvolvido pela UE – funciona como um método único de classificação para descrever o objeto dos contratos públicos, atribuindo a cada contrato um código de até nove algarismos (dos quais o nono é um algarismo de controlo), conforme o respetivo objeto [70]. Tendo em conta que, neste sistema de codificação, os dígitos mais à direita representam o objeto da relação contratual com cada vez maior especificidade, é possível obter uma classificação da relação contratual com base nos primeiros dois ou três algarismos [70]. O código CPV de cada relação contratual é obtido a partir da resposta da API do Contrato Público.

Neste contexto, admitem-se três casos distintos para os pesos das arestas, de acordo com o código CPV correspondente a cada relação contratual e o seu significado, ou seja, o tipo de serviços prestados. Estes três casos expõem-se na Tabela 15.

Peso	Descrição	Códigos CPV
0,8	A entidade contratada presta serviços com impacto potencialmente crítico na segurança da informação da entidade contratante, designadamente serviços de IT, cibersegurança ou outros que possam requerer acesso privilegiado, capacidade de administração ou possibilidade de alteração de sistemas e dados.	72******-* (Serviços de IT: consultoria, desenvolvimento de *software*, *Internet* e apoio)
0,4	A entidade contratada presta serviços com impacto potencialmente relevante na segurança da informação da entidade contratante, designadamente quando possam envolver acesso a dados sensíveis ou confidenciais, mas sem capacidades de administração ou alteração.	48******-* (Pacotes de *software* e sistemas de informação) 642*****-* (Serviços de telecomunicações) 79******-* (Serviços a empresas: direito, comercialização, consultoria, recrutamento, impressão e segurança)
0	A entidade contratada não presta serviços com impacto expectável na segurança da informação da entidade contratante, não contribuindo, por isso, para a propagação do ciber-risco no modelo considerado.	Todos os outros

Tabela 15 – Pesos das Arestas

Conforme ilustra a Tabela 15, cada aresta do grafo pode assumir um de três pesos, sendo esta determinação efetuada com base na relação contratual entre as duas entidades associadas aos vértices da aresta.

A opção por uma escala discreta de três níveis para os pesos das arestas visa modelar, de forma simples, o ciber-risco propagado entre entidades. Com esta escala, pretende-se, por um lado, evitar um nível excessivo de granularidade – que dificilmente seria suportado pela informação disponível –, mas, por outro lado, distinguir as diferentes relações contratuais com base no potencial impacto na segurança da informação da entidade contratante. Tendo em conta que o código CPV permite apenas identificar a natureza dos serviços prestados, mas não conhecer, para cada contrato, o nível efetivo de acesso concedido à entidade contratada, nem o grau de dependência operacional existente, os valores estabelecidos constituem uma estimativa da fração do ciber-risco que se admite ser propagado em cada relação contratual.

O valor mais elevado (0,8) traduz a prestação de serviços através dos quais a entidade contratada possui, por norma, maior capacidade de acesso sobre os sistemas, dados e/ou operações da entidade contratante. Por isso, admite-se que o comprometimento da primeira possa resultar num impacto significativo na segurança da informação da segunda. Este caso aplica-se, por exemplo, quando estão em causa serviços de IT ou de cibersegurança, que podem permitir a gestão remota dos equipamentos e sistemas da organização contratante, com eventuais possibilidades de acesso privilegiado, para além do acesso de leitura dos dados. Assim, considera-se que o comprometimento da entidade contratada poderia resultar numa violação crítica da confidencialidade, integridade e disponibilidade da informação da entidade contratante.

O valor intermédio (0,4) aplica-se aos serviços prestados que, devido à sua natureza, tendem a ter acesso de leitura a informação sensível da entidade contratante, mas sem o potencial disruptivo do caso anterior. A título de exemplo, enquadram-se neste âmbito os serviços de consultoria não informática, como a nível jurídico ou de recursos humanos, entre outros. Deste modo, entende-se que o comprometimento da entidade contratada poderia impactar a confidencialidade da informação da entidade contratante, não sendo expectável um impacto tão gravoso como o anterior.

Por fim, o valor mínimo (0) atribui-se às relações contratuais cujo impacto previsível na segurança da informação da entidade contratante seja mínimo ou residual. Nesta categoria, enquadram-se todas as modalidades de prestação de serviços que não encaixem nas duas anteriores.

Os valores dos pesos encontram-se entre zero e um, pois pretendem traduzir a fração do ciber-risco que é propagado da entidade contratada para a entidade contratante. Como tal, não se atribui o valor máximo de um, uma vez que a existência de uma relação contratual de qualquer tipo não implica, por si só, a propagação integral do ciber-risco da entidade contratada para a entidade contratante. Assim, a manutenção dos valores dentro deste intervalo permite representar a intensidade relativa da influência entre entidades através de uma escala comum, facilmente interpretável e compatível com os restantes cálculos efetuados pelo OSIRIS-GOV.

Tal como referido anteriormente, no caso de as mesmas duas entidades terem entre elas várias relações contratuais distintas, considera-se a relação contratual referente à aresta com o peso mais elevado.

Finalmente, após a determinação do grafo correspondente à cadeia de fornecimento da organização em análise e do cálculo dos valores dos atributos *leaks.logs*, *leaks.private.comb* e *leaks.private.general*, está recolhida toda a informação necessária para o cálculo do ciber-risco organizacional.

Assim sendo, esta informação é passada ao módulo de cálculo do ciber-risco.

4.3. Cálculo do Ciber-Risco

O módulo de cálculo do ciber-risco processa os dados recebidos pelo módulo de enriquecimento da informação e retorna uma pontuação do ciber-risco avaliado para a organização em causa, descrito pela probabilidade de ocorrência de um ciberincidente nessa organização, num dado intervalo de tempo.

Assim, o objetivo do componente inteligente deste módulo passa por prever a ocorrência de um ciberincidente numa organização, com base nos seus atributos demográficos e na exposição em fugas de dados de credenciais da organização e da sua cadeia de fornecimento. Deste modo, é necessária uma solução que seja capaz de reconhecer padrões e relações nos dados recebidos, de maneira a avaliar o ciber-risco organizacional. Como tal, recorre-se a um classificador inteligente, que processa estes atributos para prever a ocorrência de um ciberincidente. A opção por esta abordagem deve-se ao facto de a literatura recolhida ter evidenciado que as soluções de avaliação do ciber-risco tendem a assentar em algoritmos de aprendizagem automática para o efeito, dadas as suas capacidades de identificar relações entre atributos e produzir previsões com base em grandes volumes de dados.

De acordo com o estado da arte previamente mapeado, as soluções que procuram descrever o ciber-risco organizacional através de uma classificação objetiva e independente de especialistas – tal como proposto para o OSIRIS-GOV – fazem-no pela previsão de ocorrência de um ciberincidente. Por isso, o problema de aprendizagem automática em causa pode ser formulado como um problema de classificação supervisionada, no qual a variável-alvo reflete a ocorrência de um ciberincidente. Assim sendo, o sistema deve ser capaz de aprender relações entre os atributos observados em exemplos históricos previamente classificados, de modo a retornar o resultado pretendido. Efetivamente, esta abordagem encontra suporte na literatura analisada, tendo sido igualmente adotada por Liu *et al.* [37] e Sarabi *et al.* [42], cujas soluções recorrem à aprendizagem supervisionada para prever a ocorrência de futuros ciberincidentes.

Entre os algoritmos existentes para aprendizagem automática supervisionada, optou-se, para a implementação do OSIRIS-GOV, pela utilização do *random forest*. A escolha deste algoritmo deve-se à sua adequabilidade a problemas como o de avaliação do ciber-risco organizacional, devido à capacidade de capturar relações complexas entre diferentes atributos, assim como à menor propensão para aprender padrões excessivamente específicos, favorecendo a capacidade de generalização a casos futuros. De facto, foi este o algoritmo utilizado pelos autores anteriormente referidos, que também procuraram prever a ocorrência de um ciberincidente numa organização com base em informação pública, tendo obtido resultados satisfatórios [37], [42].

Após a seleção do algoritmo, procedeu-se à implementação do classificador inteligente responsável por avaliar o ciber-risco organizacional. Para tal, o algoritmo foi treinado com recurso a casos já classificados, para que fosse capaz de aprender as relações existentes entre os atributos recolhidos e a ocorrência de um ciberincidente, enquanto descritor do ciber-risco. Desta forma, o conhecimento adquirido durante a fase de treino permite que o OSIRIS-GOV preveja a ocorrência de um ciberincidente em organizações não observadas anteriormente.

A concretização desta implementação foi efetuada em *Python*, tendo sido utilizada a biblioteca *scikit-learn*[59]. Esta biblioteca disponibiliza implementações de referência para diversos algoritmos de aprendizagem automática, constituindo, por isso, uma escolha adequada para o desenvolvimento do OSIRIS-GOV.

[59] https://scikit-learn.org/

A implementação do classificador inteligente exigiu a construção de um conjunto de dados capaz de representar as características – demográficas e de exposição em fugas de dados – das organizações e a ocorrência de ciberincidentes. Este conjunto de dados constituiu a base para o processo de aprendizagem do componente inteligente do OSIRIS-GOV, permitindo treiná-lo, avaliar a sua capacidade de generalização e aferir o seu desempenho na previsão do ciber-risco organizacional. As subsecções seguintes detalham todo este processo.

4.3.1. Construção do Conjunto de Dados

Tendo em conta que o objetivo do módulo de cálculo do ciber-risco passa por prever a ocorrência de um ciberincidente numa organização, foi necessário recolher um conjunto de organizações vítimas de ciberincidentes e um conjunto de organizações que não foram vítimas de ciberincidentes.

O treino e teste com este conjunto de dados permite capacitar o classificador inteligente para distinguir entre organizações que sofreram um ciberincidente e organizações que nunca sofreram um ciberincidente registado publicamente na VCDB. Esta abordagem é idêntica à adotada pelos autores anteriores, que também utilizaram um algoritmo *random forest* para prever a ocorrência de ciberincidentes nas organizações [37], [42].

Tanto para as vítimas, quanto para as não vítimas, todas as organizações deveriam estar acompanhadas da respetiva informação demográfica – como o país, o número de colaboradores e a indústria ou o setor de atividade – e de informação relativa à respetiva exposição de credenciais em fugas de dados. A necessidade de obter esta informação prende-se com o proposto para o OSIRIS-GOV, ou seja, aferir o ciber-risco de uma organização, a partir de informação pública, nomeadamente, dos atributos demográficos e da exposição de credenciais em fugas de dados.

No entanto, na literatura previamente recolhida, não foi encontrado nenhum conjunto de dados público que pudesse, sem qualquer processamento adicional, ser diretamente aplicável ao pretendido pelo OSIRIS-GOV. Assim sendo, o conjunto de dados adequado à arquitetura proposta teve de ser construído como parte integrante do trabalho realizado.

A Figura 6 ilustra o processo de construção deste conjunto de dados.

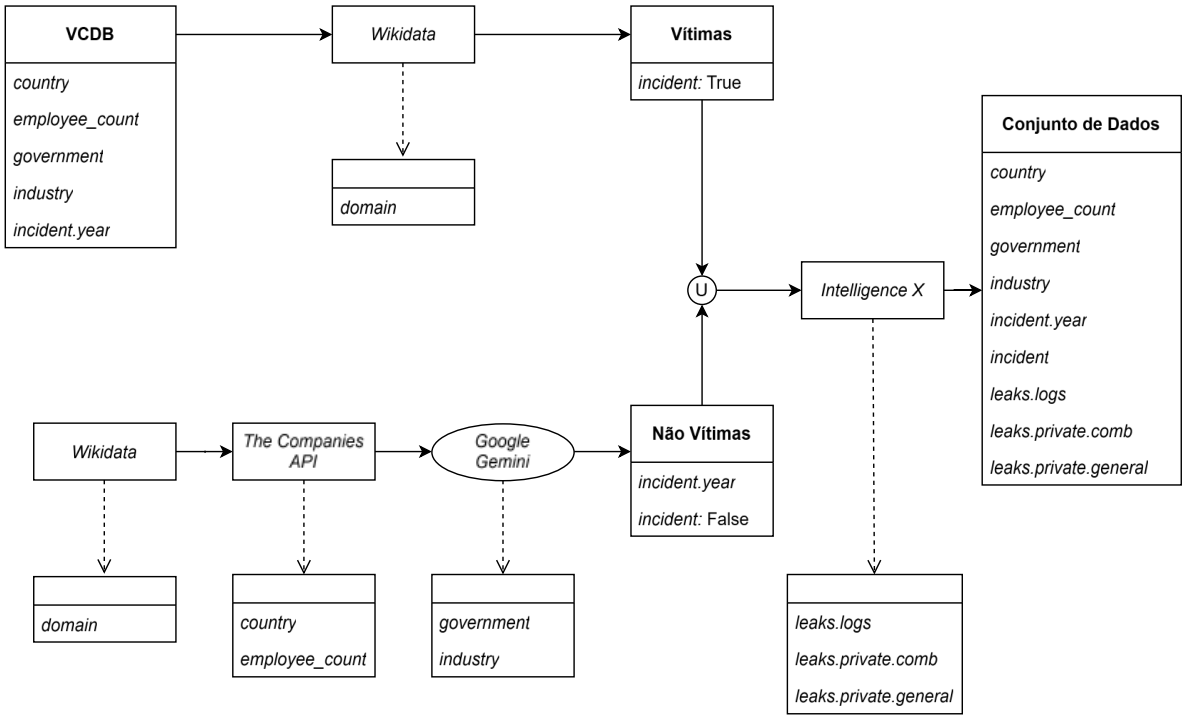

Figura 6 – Construção do Conjunto de Dados

Conforme evidencia a Figura 6, a construção do conjunto de dados seguiu dois fluxos principais. O fluxo superior representa a recolha de dados das organizações vítimas de ciberincidentes (num conjunto designado “Vítimas”), enquanto o fluxo inferior traduz o processo análogo, para os dados das organizações que não foram vítimas de ciberincidentes (num conjunto denominado “Não Vítimas”).

Seguindo o fluxo ilustrado na parte superior do diagrama, a recolha das organizações vítimas de ciberincidentes partiu da VCDB, enquanto base de dados – identificada no levantamento do estado da arte – que contém organizações que sofreram um ciberincidente tornado público. Esta base de dados armazena, num formato estruturado, um conjunto de atributos demográficos sobre mais de 8000 organizações vítimas de ciberincidentes, bem como o ano em que sofreram o ciberincidente que motivou o registo na VCDB [71]. Assim sendo, foram recolhidos todos os registos da VCDB, ou seja, todas as organizações vítimas de um ciberincidente tornado público e reportado nesta base de dados, incluindo-se unicamente os atributos referentes ao país (*country*), número de colaboradores (*employee_count*), pertença ao setor da administração pública (*government*) e indústria/setor de atividade (*industry*), bem como o ano no qual foi sofrido o ciberincidente (*incident.year*).

Como, para estas entidades, era necessário obter informações sobre as fugas de dados com exposição de credenciais associadas ao respetivo domínio, teve de ser identificado o domínio de cada organização. Para esse efeito, foi consultada a API da *Wikidata*[60], enviando o valor do atributo *victim_id* de cada registo da VCDB para, em caso de existência de um registo correspondente na base de dados da *Wikidata*, receber como resposta o domínio correspondente. Desta forma, foi possível obter o domínio de cerca de metade das entidades registadas na VCDB, armazenando o resultado de cada uma no atributo *domain* da linha correspondente.

A partir destas entidades, foi criado o conjunto de organizações vítimas de ciberincidentes. Para isso, foi acrescentado a todas estas organizações um atributo booleano *incident*, sempre com valor verdadeiro (*True*), que corresponde à variável-alvo ou etiqueta (*label*) do conjunto de dados. Assim, o subconjunto "Vítimas" representa as organizações que foram, efetivamente, vítimas de um ciberincidente registado na VCDB.

O fluxo inferior do diagrama descreve visualmente o processo de construção do conjunto de não vítimas, igualmente necessário para o desenvolvimento do classificador. Em primeiro lugar, foram recolhidos domínios da base de dados da *Wikidata*, garantindo que estes domínios não se encontravam no conjunto de vítimas identificadas anteriormente. Esta lista de domínios – armazenados no atributo *domain* – constituiu o ponto de partida para a construção do conjunto de não vítimas.

Posteriormente, cada um destes domínios foi enviado à *The Companies API*[61], para recolha de informação demográfica sobre a respetiva organização. Esta API devolveu, para cada domínio registado na sua base de dados, o país (*country*) e o número de colaboradores (*employee_count*) da organização associada, bem como uma lista de palavras-chave descritivas da área de atividade da mesma. Em seguida, estas palavras-chave foram submetidas a um modelo da família *Google Gemini*[62], de modo a obter a respetiva indústria/setor de atividade e a indicação de pertença (ou não) ao setor da administração pública. Em particular, o modelo selecionado foi o *Large Language Model* (LLM) *Gemini 3.1 Flash-Lite*, por ser aquele que permitia, de forma gratuita via API, processar mais pedidos num mesmo intervalo de tempo. Assim, este LLM retornou os valores dos atributos *government* e *industry* para cada organização correspondente aos domínios submetidos.

[60] https://www.wikidata.org/
[61] https://www.thecompaniesapi.com/
[62] https://gemini.google.com/

Deste modo, foram recolhidos os mesmos atributos provenientes da VCDB, mas, desta feita, para um conjunto de organizações que não sofreram ciberincidentes registados na VCDB. Tal como anteriormente, foi adicionado um atributo booleano *incident* a todas estas organizações, mas, neste caso, com valor falso (*False*), sendo este atributo a variável-alvo/etiqueta do conjunto de dados. Foi ainda atribuído um valor sintético ao atributo *incident.year*, seguindo, aproximadamente, a mesma distribuição dos valores deste atributo no conjunto de organizações vítimas, para evitar potenciais enviesamentos do classificador. Assim, criou-se o subconjunto "Não Vítimas".

Com isto, existia um conjunto de vítimas e um conjunto de não vítimas, ambos com os mesmos atributos: domínio (*domain*), país (*country*), número de colaboradores (*employee_count*), indicação de pertença ao setor da administração pública (*government*), indústria/setor de atividade (*industry*), ano de ciberincidente (real nas vítimas e sintético nas não vítimas) (*incident.year*) e indicação de presença (ou não) na VCDB, enquanto vítima de um ciberincidente registado (*incident*).

A partir daqui, estes dois conjuntos foram unidos, para pesquisa no *Intelligence X*. Seguindo a arquitetura pretendida para o OSIRIS-GOV, foram pesquisadas, para cada organização, as fugas de dados com credenciais expostas associadas ao respetivo domínio. Para esse efeito, a pesquisa foi efetuada nos *buckets leaks.logs*, *leaks.private.comb* e *leaks.private.general* do *Intelligence X*, através do domínio de cada organização. Tendo em conta o objetivo pretendido para o OSIRIS-GOV, ou seja, prever a ocorrência de um ciberincidente futuro com base nos atributos demográficos e na exposição de credenciais em fugas de dados de cada organização, o intervalo de tempo delimitado para pesquisa consistiu nos cinco anos imediatamente anteriores ao ano do ciberincidente registado. Por exemplo, para uma entidade com um ciberincidente registado em 2025, foram pesquisadas as fugas de dados com credenciais expostas entre 2020 e 2024 (inclusive). A frequência absoluta de fugas de dados associadas a cada organização foi armazenada num atributo com o nome correspondente ao *bucket* de origem dos dados, povoando as colunas *leaks.logs*, *leaks.private.comb* e *leaks.private.general.*

Após a pesquisa, o atributo *domain* foi removido do conjunto de dados. Esta decisão foi tomada por se considerar que o domínio da organização assume apenas uma natureza identificadora, logo, sem capacidade de aumentar o valor preditivo do componente inteligente. Por isso, este atributo foi eliminado na tentativa de maximizar as capacidades de previsão e de generalização do OSIRIS-GOV, tendo sido apenas necessário para a realização das pesquisas anteriores.

Assim sendo, todo este processo resultou na criação de um conjunto de dados com atributos demográficos e atributos indicadores da exposição em fugas de dados de credenciais corporativas de cada organização.

A Tabela 16 sintetiza os atributos deste conjunto de dados.

Atributo	Tipo	Descrição
country	Categórico	País de atividade.
employee_count	Categórico	Número de colaboradores (inferior ou igual a mil, superior a mil, ou desconhecido).
government	Categórico	Indicação de pertença ao setor da administração pública.
industry	Categórico	Setor de atividade.
incident.year	Numérico	Ano de referência (ano real de ciberincidente nas vítimas, ou sintético nas não vítimas).
incident (Variável-Alvo)	Categórico	Indicação de ocorrência efetiva de um ciberincidente no ano de referência.
leaks.logs	Numérico	Frequência absoluta de fugas de dados provenientes de *malware infostealers,* no intervalo de tempo considerado.
leaks.private.comb	Numérico	Frequência absoluta de fugas de dados provenientes de listas e compilações, no intervalo de tempo considerado.
leaks.private.general	Numérico	Frequência absoluta de fugas de dados não enquadradas nos outros *buckets,* no intervalo de tempo considerado.

Tabela 16 – Atributos do Conjunto de Dados

Conforme sinalizado na Tabela 16, o atributo *incident* é a variável-alvo. Este atributo categórico assume um valor booleano, isto é, verdadeiro ou falso, indicando se a organização sofreu, efetivamente, um ciberincidente registado na VCDB, no ano considerado como referência (*incident.year*). Assim, o problema foi formulado como uma tarefa de classificação binária probabilística, na qual se pretende estimar a probabilidade de ocorrência de um ciberincidente com base nos restantes atributos do conjunto de dados.

Tendo já recolhido os valores correspondentes a todos os atributos, foi efetuada uma breve análise da qualidade dos dados.

A pesquisa de fugas de dados no *Intelligence X* revelou que, a partir de um determinado ano, os resultados obtidos eram extremamente limitados, com um elevado número de organizações que não tinham quaisquer fugas de dados associadas.

A Tabela 17 exemplifica os resultados obtidos para o subconjunto de vítimas pesquisadas no *Intelligence X*.

Ano	**Vítimas**	**Vítimas com Fugas de Dados no *Intelligence X***	
2012	433	177	(41%)
2013	678	320	(47%)
2014	378	160	(42%)
2015	409	179	(44%)
2016	410	194	(47%)
2017	276	110	(40%)
2018	195	106	(54%)
2019	155	92	(59%)
2020	190	140	(74%)
2021	96	79	(82%)
2022	66	64	(97%)
2023	598	591	(99%)
2024	59	57	(97%)
2025	92	90	(98%)

Tabela 17 – Vítimas com Fugas de Dados no *Intelligence X*

Tal como é visível na Tabela 17, só a partir de 2018 é que o *Intelligence X* consegue captar fugas de dados associadas a pelo menos metade das organizações vítimas pesquisadas. Até lá, a maioria das vítimas procuradas nesta fonte de informação retornou zero fugas de dados, em todos os *buckets*. Como tal, este facto levou a que se optasse por incluir no conjunto de dados apenas as organizações com um ano de referência a partir de 2018 (inclusive), ou seja, desde 2018 até 2025.

Após este filtro, o conjunto de dados – constituído por 3132 organizações, das quais 1451 são vítimas e 1681 não vítimas – ficou pronto para a aplicação do algoritmo *random forest*.

4.3.2. *Random Forest*

Concluída a construção do conjunto de dados, foi, conforme já referido, utilizado o algoritmo *random forest* para implementar o componente inteligente do OSIRIS-GOV, pelas razões previamente explicitadas.

O algoritmo *random forest* é constituído por um conjunto de árvores de decisão. Uma árvore de decisão é um algoritmo de aprendizagem supervisionada, para tarefas de classificação ou de regressão, cujo processamento segue uma estrutura hierárquica semelhante a uma árvore, desde a raiz até às folhas [72].

A Figura 7 exemplifica a estrutura de uma árvore de decisão.

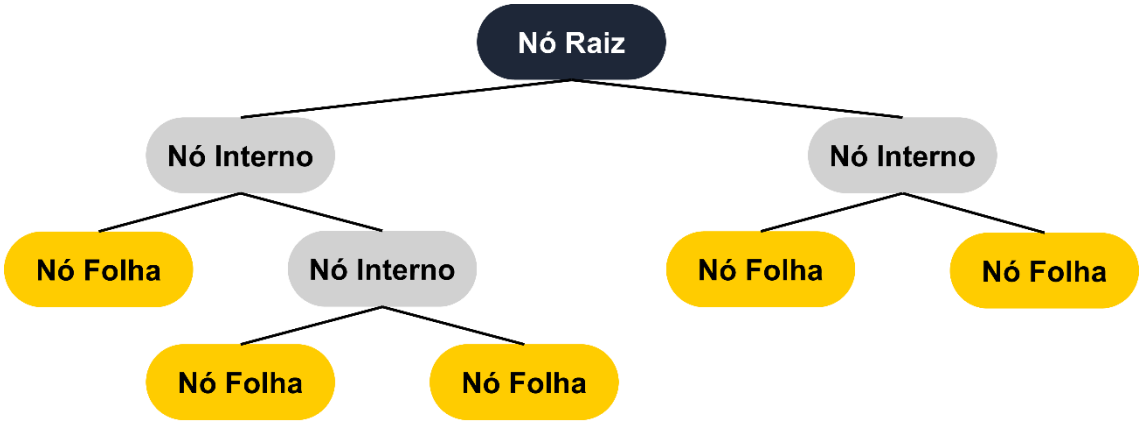

Figura 7 – Árvore de Decisão

Tal como transmitido visualmente na Figura 7, uma árvore de decisão é constituída por três tipos de nós: o nó raiz, os nós internos (ou nós de decisão) e os nós folha.

O nó raiz é o nó inicial da árvore, no qual é introduzido o conjunto completo de dados. Juntamente com os nós internos, todos estes aplicam determinadas condições com base nos valores dos atributos do conjunto de dados, na tentativa de formar subconjuntos cada vez mais homogéneos em relação à variável-alvo do problema de classificação ou regressão. Para isso, em cada decisão, o algoritmo tenta selecionar o atributo que melhor separa os dados, relativamente à variável-alvo. Assim, o conjunto de dados vai sendo sucessivamente processado e separado pelo algoritmo, até chegar aos nós folha, que representam o valor previsto pelo algoritmo.

Um classificador assente numa árvore de decisão tem a vantagem de ser facilmente explicável e interpretável, dada a simplicidade e a possibilidade de visualização gráfica do processo de tomada de decisão. No entanto, este algoritmo apresenta como uma das principais limitações a tendência para o *overfitting*, ou seja, uma baixa capacidade de generalização, devida a uma memorização excessiva dos padrões do conjunto de treino. A par disto, uma árvore de decisão é, frequentemente, considerada como um algoritmo instável, visto que ligeiras alterações no conjunto de dados tendem a provocar a criação de árvores substancialmente diferentes. Por isso, surge a necessidade de aplicar algoritmos de aprendizagem conjunta, que combinam múltiplas árvores de decisão, como é o caso do *random forest*.

O *random forest* é um algoritmo de aprendizagem supervisionada constituído, tal como o nome indica, por um conjunto de árvores de decisão, ou seja, uma floresta de árvores de decisão. A ideia subjacente à utilização deste algoritmo passa pela tentativa de minimização do risco de *overfitting* e pela redução da instabilidade associada às árvores de decisão.

Para isso, o *random forest* começa por atribuir um conjunto aleatório dos dados a cada árvore de decisão. Desta forma, garante-se que cada árvore é treinada com dados ligeiramente diferentes. Além disto, em cada divisão efetuada pelas árvores de decisão, só é considerado um subconjunto aleatório dos atributos disponíveis, na tentativa de aumentar a diversidade entre as árvores e reduzir potenciais correlações entre as previsões de cada uma. Com isto, cada árvore de decisão produz o seu resultado para o problema em causa, sendo todos os resultados combinados pelo algoritmo *random forest* consoante o tipo de problema.

A Figura 8 ilustra esta organização interna do algoritmo *random forest*.

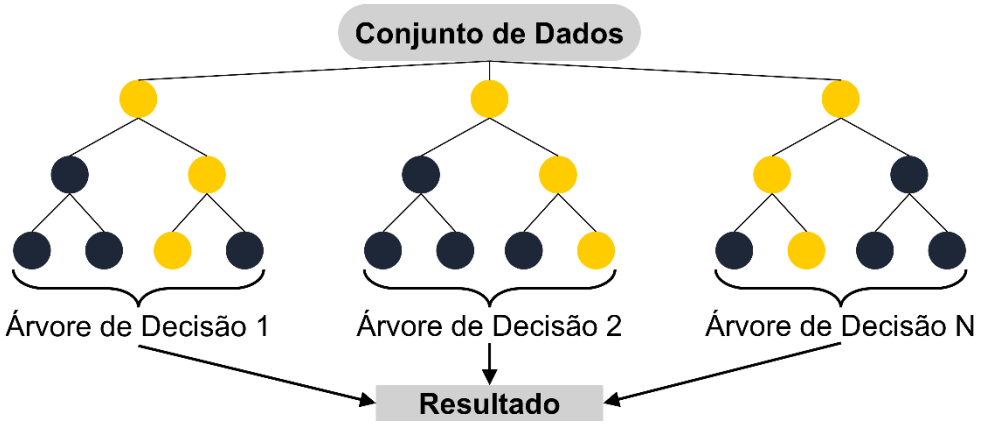

Figura 8 – *Random Forest*

Conforme evidencia a Figura 8, o *random forest* combina os resultados intermédios provenientes das várias árvores de decisão para obter um resultado global. No caso de problemas de classificação, o valor de saída do algoritmo tende a ser o resultado da votação maioritária de todas as árvores de decisão. No caso de problemas de regressão, é utilizada uma medida de agregação para as previsões das diferentes árvores de decisão, que, frequentemente, consiste na média das mesmas.

Em comparação com uma árvore de decisão, o algoritmo *random forest* tende a ter uma melhor capacidade de generalização, com maior resistência a ruído e a valores extremos, reduzindo o risco de *overfitting*. Esta melhoria global no desempenho tem associada uma maior dificuldade na interpretação ou explicação do processo de tomada de decisão por parte do algoritmo, dada a menor simplicidade do mesmo, mas esta abordagem tende a ser preferida pelos seus melhores resultados.

Por tudo isto, juntamente com as evidências de bons resultados obtidos pela utilização deste algoritmo na literatura científica recolhida, o *random forest* foi a opção tomada para ser o algoritmo do componente inteligente do OSIRIS-GOV.

4.3.3. Treino e Teste

O algoritmo escolhido para o OSIRIS-GOV teve de passar por um processo de treino e teste, com o conjunto de dados previamente construído.

Conforme definido anteriormente, o atributo *incident* foi considerado como a variável-alvo do classificador, enquanto os restantes atributos foram utilizados como variáveis de entrada para prever a probabilidade de ocorrência de um ciberincidente numa organização.

Sendo o objetivo a previsão da probabilidade de ocorrência de um ciberincidente numa organização num determinado ano de referência, este é um problema de classificação binária probabilística. Como tal, foi utilizado um algoritmo *random forest* preparado para um problema de classificação (*RandomForestClassifier*[63], da biblioteca *scikit-learn*), a partir do qual foi possível – ao invocar a função *predict_proba()* da mesma biblioteca – obter a probabilidade prevista para a ocorrência de um ciberincidente. Esta probabilidade corresponde à média das probabilidades estimadas pelas árvores de decisão do *random forest*, sendo que, em cada árvore de decisão, a probabilidade de uma classe é calculada a partir da proporção de amostras dessa classe na folha alcançada.

Como os atributos *country*, *employee_count*, *government* e *industry* do conjunto de dados são categóricos, o algoritmo implementado exige que sejam codificados de forma a permitir o seu processamento. Para isso, foi utilizado *one-hot encoding*. Esta técnica consiste em converter um atributo que pode assumir N valores possíveis em $N-1$ novos atributos binários, representando as mesmas categorias, mas de forma numérica. Os atributos correspondentes à informação sobre fugas de dados, por serem numéricos, foram mantidos sem qualquer transformação.

A partir daqui, o conjunto de dados foi dividido em subconjuntos de treino e teste. De acordo com a abordagem aplicada na literatura por Liu *et al.* [37] e Sarabi *et al.* [42], os dados foram separados de forma cronológica para treino e teste. Esta estratégia foi considerada a mais adequada para este tipo de problemas, nos quais o objetivo do classificador passa por prever a probabilidade de uma ocorrência futura com base nos valores observados no passado. Assim, a informação referente a um determinado intervalo de anos foi utilizada como treino, sendo o teste realizado nos dados referentes ao ano seguinte. Para isso, foram efetuadas todas as partições possíveis de anos consecutivos desde 2018 até 2025. A título de exemplo, o teste do classificador no ano de 2022 correspondeu à realização do treino nos valores compreendidos entre 2018 e 2021. O atributo *incident.year* foi exclusivamente utilizado para esta separação temporal, tendo sido removido após a realização da mesma.

[63] https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

Adicionalmente, a conversão das probabilidades previstas pelo classificador em classificações binárias requer a definição de um limiar de discriminação. Este limiar corresponde à probabilidade a partir da qual uma observação é classificada como positiva. Embora, por omissão, o valor predefinido para o limiar de discriminação seja 0,5, esta nem sempre é a opção mais adequada, uma vez que diferentes valores podem influenciar o desempenho do classificador nas diversas métricas de avaliação.

Assim, antes de avaliar o classificador em cada subconjunto de dados de teste, foi utilizado o subconjunto correspondente ao ano imediatamente anterior ao ano de teste para determinar o limiar de discriminação. Para tal, foi selecionado o valor que maximizava o desempenho do classificador nas métricas de avaliação consideradas, no subconjunto relativo ao ano anterior ao ano de teste. Após a determinação do limiar de discriminação, o classificador foi novamente treinado utilizando todos os dados disponíveis até ao ano anterior ao ano de teste, inclusive, para ser testado e avaliado no subconjunto correspondente ao ano de teste.

Seguindo a mesma abordagem dos autores elencados anteriormente, o *random forest* foi instanciado com 50 classificadores, isto é, 50 árvores de decisão. Posto isto, foi treinado e testado nos sucessivos intervalos de anos de acordo com a estratégia delineada, tendo sido avaliado posteriormente.

4.3.4. Avaliação

De maneira a validar o desempenho do classificador inteligente, foram calculadas cinco métricas frequentemente utilizadas para avaliação: exatidão (*accuracy*), precisão (*precision*), *recall*, pontuação F_1 (F_1-*score*) e a *Area Under the Curve* (AUC), isto é, a área sob a curva *Receiver Operating Characteristic* (ROC).

De acordo com a metodologia de treino e teste explicitada anteriormente, o teste num determinado ano correspondeu ao treino no subconjunto de dados relativo a todos os anos anteriores, a partir de 2018, inclusive. Como tal, o primeiro ano de teste é 2020, com o treino respetivo em 2018, a validação para definir o limiar de discriminação em 2019 e o novo treino em 2018 e 2019. Por sua vez, 2025 é o último ano de teste, com o treino inicial no intervalo de tempo desde 2018 até 2023 e a definição do limiar de discriminação no ano de 2024, incidindo o treino final sobre os dados de 2018 a 2024. A métrica de avaliação do desempenho que se pretendeu maximizar com o ajuste do limiar de discriminação foi a pontuação F_1, por ser aquela que traduz, simultaneamente, o equilíbrio entre a precisão e o *recall*.

Para efeitos de avaliação, considera-se como verdadeiro positivo (VP) uma organização corretamente classificada como vítima de um ciberincidente no ano de referência. De forma análoga, um verdadeiro negativo (VN) corresponde a uma organização corretamente classificada como não vítima. Um falso positivo (FP) representa uma organização incorretamente classificada como vítima, enquanto um falso negativo (FN) é uma organização incorretamente classificada como não vítima.

Para validação da abordagem utilizada e aferição efetiva do valor preditivo adicional aportado pelos atributos referentes às fugas de dados, os valores obtidos para as cinco métricas calculadas foram comparados com os valores correspondentes, mas relativos a um classificador que apenas considerasse os atributos demográficos de cada organização, ou seja, sem as colunas *leaks.logs*, *leaks.private.comb* e *leaks.private.general*. Este termo de comparação designa-se "Atributos Demográficos", enquanto o componente inteligente utilizado pelo OSIRIS-GOV denomina-se "Atributos Demográficos e Fugas de Dados". O objetivo desta comparação passa por perceber se a informação relativa à exposição de cada organização em fugas de dados aumenta, efetivamente, a capacidade preditiva do classificador, comparando-o com abordagens similares às identificadas no levantamento do estado da arte.

Assim, calcularam-se a exatidão, precisão, *recall*, pontuação F_1 e AUC de cada um destes dois classificadores, seguindo, em ambos os casos, a estratégia delineada anteriormente.

A exatidão corresponde à proporção de previsões corretas, sendo dada pelo cálculo da expressão da Equação (5).

$$Exatidão = \frac{VP + VN}{VP + VN + FP + FN} \tag{5}$$

A Tabela 18 apresenta os resultados obtidos para esta métrica.

Ano de Teste / Classificador	2020	2021	2022	2023	2024	2025
Atributos Demográficos	0,78	0,78	0,84	0,79	0,84	0,78
Atributos Demográficos e Fugas de Dados	0,79	0,71	0,84	0,87	0,80	0,82

Tabela 18 – Exatidão

Os resultados apresentados na Tabela 18 mostram um desempenho muito similar entre os dois classificadores. Como tal, embora se observe uma ligeira melhoria do classificador que contempla as fugas de dados em relação ao que considera apenas os atributos demográficos conforme aumenta o tamanho do subconjunto de dados de treino, isto é, quanto mais recente é o ano de teste, as diferenças entre os resultados obtidos para esta métrica não são muito expressivas.

A precisão corresponde à proporção de previsões positivas que estão corretas, isto é, a frequência relativa de casos previstos como ciberincidentes que são, efetivamente, verdadeiros ciberincidentes.

Assim sendo, é calculada pela expressão da Equação (6).

$$Precisão = \frac{VP}{VP + FP} \tag{6}$$

A Tabela 19 mostra os valores calculados para a precisão.

Classificador \ Ano de Teste	2020	2021	2022	2023	2024	2025
Atributos Demográficos	0,77	0,75	0,82	0,69	0,98	0,84
Atributos Demográficos e Fugas de Dados	0,79	0,66	0,79	0,80	0,92	0,82

Tabela 19 – Precisão

Tal como anteriormente, também os resultados da Tabela 19 não evidenciam diferenças notórias no desempenho dos dois classificadores, relativamente à precisão. Os valores obtidos para os diferentes anos de teste sofrem de alguma variabilidade de ano para ano, não sendo possível, a partir da análise isolada destes resultados, aferir a melhoria da capacidade preditiva com a introdução da informação sobre fugas de dados.

O *recall* é a proporção de exemplos positivos que são capturados pelo classificador, ou seja, a frequência relativa de verdadeiros ciberincidentes identificados. Na prática, corresponde ao rácio de verdadeiros positivos. A expressão para o cálculo desta métrica é a da Equação (7).

$$Recall = \frac{VP}{VP + FN} \quad (7)$$

A Tabela 20 elenca os resultados de *recall*.

Ano de Teste / Classificador	2020	2021	2022	2023	2024	2025
Atributos Demográficos	0,74	0,81	0,83	0,98	0,68	0,64
Atributos Demográficos e Fugas de Dados	0,73	0,80	0,86	0,96	0,79	0,78

Tabela 20 – *Recall*

Os resultados obtidos para o *recall* – expostos na Tabela 20 – começam a evidenciar uma tendência mais relevante de melhoria do classificador que considera as fugas de dados, em relação ao que se cinge aos atributos demográficos das organizações. Em particular, ao aumentar o número de casos de treino, ou seja, com um intervalo de anos para treino cada vez maior, verifica-se um desempenho consistentemente superior do classificador "Atributos Demográficos e Fugas de Dados". Esta diferença é particularmente importante dado que o *recall* mede a capacidade para captar organizações que sofreram, efetivamente, um ciberincidente, o que sugere que as variáveis associadas às fugas de dados aumentam a capacidade na identificação destas organizações vítimas, reduzindo a proporção de ciberincidentes não detetados.

Dada a existência de um compromisso (*tradeoff*) entre a precisão e o *recall*, na medida em que o aumento do valor de uma das métricas pode tender a reduzir o valor da outra, existe a métrica F (F-*measure*) ou pontuação F (F-*score*), que combina a precisão e o *recall*. A expressão geral de cálculo desta métrica é a da Equação (8).

$$F_\beta = \frac{(1 + \beta^2) \times Precisão \times Recall}{\beta^2 \times Precisão + Recall} \quad (8)$$

Na expressão de cálculo de F_β, o parâmetro β controla a importância relativa da precisão e do *recall*, sendo que uma diminuição do valor de β leva à atribuição de menos peso ao valor de *recall*, enquanto um aumento do valor de β implica a diminuição da importância atribuída à precisão.

Quando o valor de β é 1, F_β designa-se F_1 e corresponde à média harmónica entre os valores de precisão e *recall*, sendo calculada pela expressão da Equação (9).

$$F_1 = \frac{2 \times Precisão \times Recall}{Precisão + Recall} \quad (9)$$

A Tabela 21 lista os resultados obtidos para a métrica F_1.

Classificador \ Ano de Teste	2020	2021	2022	2023	2024	2025
Atributos Demográficos	0,75	0,78	0,83	0,81	0,80	0,73
Atributos Demográficos e Fugas de Dados	0,76	0,73	0,83	0,87	0,85	0,80

Tabela 21 – F_1

Como seria expectável, dados os valores previamente observados para a precisão e para o *recall*, os resultados da Tabela 21 traduzem uma melhoria global do classificador que inclui as fugas de dados na análise, quanto à pontuação F_1. Estes resultados são relevantes porque demonstram que a melhoria anteriormente evidenciada pelo classificador "Atributos Demográficos e Fugas de Dados" quanto à métrica *recall* não ocorre à custa de uma degradação significativa da precisão. Assim, com o crescimento do número de dados de treino, associado ao aumento do ano de teste, os resultados de F_1 sustentam que a inclusão das fugas de dados nos atributos a fornecer tende a melhorar o equilíbrio global entre a precisão e o *recall*, ou seja, entre identificar corretamente organizações vítimas e evitar classificações positivas incorretas.

Finalmente, a curva ROC é a visualização gráfica que ilustra o compromisso entre o rácio de verdadeiros positivos e o rácio de falsos positivos, conforme varia o limiar de discriminação entre as duas classes.

O rácio de verdadeiros positivos é a métrica calculada pela expressão da Equação (7) (*recall*), enquanto o rácio de falsos positivos é a proporção de casos negativos classificados, incorretamente, como positivos, calculada pela expressão da Equação (10).

$$Rácio\ de\ Falsos\ Positivos = \frac{FP}{FP + VN} \quad (10)$$

Exemplificativamente, o ponto $(0,0)$ na curva ROC corresponde à classificação de todos os casos como negativos, enquanto o ponto $(1,1)$ corresponde à classificação de todos os casos como positivos. Deste modo, o ponto que se pretende atingir é o $(0,1)$, correspondendo a um classificador ideal, na medida em que distingue perfeitamente ambas as classes.

A AUC é, tal como o nome indica, a área sob a curva ROC, que indica quão bom é o classificador a distinguir as duas classes (positiva e negativa) do conjunto de dados. Assim, um classificador perfeito tem uma AUC de 1, enquanto um classificador aleatório tem, em média, uma AUC de 0,5.

Os resultados obtidos para a AUC – relativos a ambos os classificadores – apresentam-se na Tabela 22.

Ano de Teste / Classificador	2020	2021	2022	2023	2024	2025
Atributos Demográficos	0,84	0,84	0,90	0,94	0,91	0,85
Atributos Demográficos e Fugas de Dados	0,86	0,86	0,91	0,96	0,94	0,90

Tabela 22 – AUC

Os resultados da Tabela 22 mostram, de forma clara e consistente, o melhor desempenho obtido pelo classificador “Atributos Demográficos e Fugas de Dados”. Em todos os anos de teste, este classificador obtém uma AUC superior ao classificador que se limita aos atributos demográficos, sendo esta melhoria cada vez mais visível com o aumentar do ano de teste e, consequentemente, do tamanho do subconjunto de treino. Deste modo, os resultados obtidos para a AUC indicam que as variáveis relativas às fugas de dados acrescentam informação útil para a capacidade preditiva e discriminatória do classificador.

Globalmente, embora os resultados obtidos para a exatidão e precisão não apresentem diferenças relevantes, os valores obtidos para *recall*, pontuação F_1 e AUC sugerem que a inclusão de atributos relativos a fugas de dados acrescenta valor preditivo ao classificador desenvolvido. Assim sendo, esta conclusão sustenta a escolha do classificador "Atributos Demográficos e Fugas de Dados" para ser utilizado pela solução OSIRIS-GOV, tendo em conta a sua melhor capacidade de previsão e classificação.

4.4. Validação

A implementação adotada para o OSIRIS-GOV assegurou o cumprimento de todos os requisitos anteriormente estabelecidos, tanto funcionais, quanto não funcionais.

Em primeiro lugar, observem-se os requisitos funcionais. O cumprimento do RF1 é atingido, dado que o NIPC de cada organização atua como o seu identificador único, introduzido manualmente pelo utilizador como valor de entrada no OSIRIS-GOV, para consulta na base de dados de organizações. A partir deste identificador, são recolhidos da PORDATA os atributos demográficos da entidade introduzida, pelo que é cumprido o RF2. De seguida, também são recolhidas, mas do *Intelligence X*, informações relativas às fugas de dados com credenciais corporativas expostas da organização em análise, indo ao encontro do RF3. No que concerne à cadeia de fornecimento da organização em análise, conforme contemplado no RF4, o mapeamento da mesma é efetuado com recurso ao Contrato Público e ao NIF.PT, sendo todas as entidades identificadas incluídas na avaliação do ciber-risco. Em conformidade com o RF5, a diferenciação entre o contributo de cada entidade para a avaliação do ciber-risco é realizada com base no código CPV de cada relação contratual. Para todas estas entidades da cadeia de fornecimento, também são consultadas, a partir do *Intelligence X*, as fugas de dados com credenciais expostas, de modo a satisfazer o RF6. Toda a informação recolhida – incluindo os atributos demográficos da organização em análise e as fugas de dados da própria organização, bem como da respetiva cadeia de fornecimento – é tida em consideração na avaliação do ciber-risco, de acordo com o RF7. Por fim, o OSIRIS-GOV retorna como resultado a probabilidade de ocorrência de um ciberincidente na organização em causa, enquanto pontuação descritora do ciber-risco avaliado, cumprindo com o RF8.

Analogamente, as mesmas considerações podem ser tecidas face aos requisitos não funcionais. O RNF1 é cumprido, visto que as fontes de informação consultadas para a recolha de informação são a PORDATA, o *Intelligence X*, o NIF.PT e o Contrato Público, logo, de acesso disponível ao público. Relativamente ao RNF2 (Desempenho), os testes efetuados demonstraram tempos de execução consistentemente inferiores a cinco minutos, desde a introdução do valor de entrada por parte do utilizador, até ao retorno do resultado pelo OSIRIS-GOV, em cumprimento do estabelecido. Por sua vez, o RNF3 (Disponibilidade) também é satisfeito no que concerne à execução local da solução, embora a recolha de dados esteja condicionada à disponibilidade das fontes de informação externas. Quanto ao RNF4 (Privacidade), os únicos dados recolhidos pelo OSIRIS-GOV são os atributos demográficos da organização em análise e informação estatística sobre as fugas de dados da organização e dos seus terceiros, fornecedores e prestadores de serviços, logo, não são recolhidos quaisquer dados sensíveis, indo ao encontro deste requisito. O RNF5 (Usabilidade) também é satisfeito, visto que a interação com o utilizador é realizada de forma simples, através do terminal, decorrendo o funcionamento interno da solução de forma transparente para o utilizador. Internamente, o código-fonte da solução encontra-se devidamente documentado, de maneira a cumprir o RNF6 (Manutenibilidade). Por último, conforme já explicitado, a arquitetura do sistema é modular e facilmente extensível a novas fontes de informação, pelo que o RNF7 (Extensibilidade) também é satisfeito.

Deste modo, a implementação da arquitetura delineada para o OSIRIS-GOV possibilitou o cumprimento de todos os requisitos, quer dos oito funcionais, quer dos sete não funcionais.

5. Caso de Estudo

Este capítulo apresenta um caso de estudo da aplicação do OSIRIS-GOV ao contexto da administração pública local portuguesa, destacando os resultados obtidos pela solução proposta para um conjunto de câmaras municipais e comparando o impacto da inclusão ou exclusão das respetivas cadeias de fornecimento.

5.1. Enquadramento

De maneira a averiguar a viabilidade da aplicação do OSIRIS-GOV a um contexto real, foi estudado o caso da administração pública local portuguesa, ou seja, das câmaras municipais nacionais.

A escolha deste setor para a realização deste caso de estudo deve-se, essencialmente, a dois fatores que o tornam particularmente relevante para a análise. Por um lado, conforme aludido inicialmente, as tendências mais recentes mostram que o setor da administração pública tem vindo a ser dos principais impactados pelas ciberameaças atuais. Por outro lado, sendo este setor altamente regulado e com requisitos de transparência, existe tendencialmente mais informação disponível publicamente e útil para a aferição do ciber-risco, em comparação com o setor privado. Em particular, a informação relativa aos contratos públicos celebrados pelas autarquias permite identificar, de forma não intrusiva, os terceiros, fornecedores e prestadores de serviços de cada uma destas entidades, ou seja, mapear a respetiva cadeia de fornecimento.

Assim sendo, para efeitos do presente caso de estudo, foram selecionadas 24 câmaras municipais, sobre as quais foi aplicada a proposta de solução OSIRIS-GOV. Tendo em vista a tentativa de assegurar uma maior representatividade da amostra de câmaras municipais escolhidas, esta seleção incidiu sobre 12 câmaras municipais com mais de mil colaboradores – Albufeira, Almada, Faro, Leiria, Lisboa, Loures, Porto, Seixal, Sintra, Viana do Castelo, Vila do Conde e Vila Nova de Gaia – e 12 câmaras municipais com menos de mil colaboradores – Aveiro, Borba, Bragança, Caminha, Castelo de Paiva, Cinfães, Mirandela, Paços de Ferreira, Ponta Delgada, Santarém, Santo Tirso e Vizela –, de acordo com os dados da PORDATA. Note-se que esta amostra não pretende representar estatisticamente todas as câmaras municipais portuguesas, mas apenas permitir a realização de uma análise exploratória no âmbito do presente caso de estudo.

Para cada uma destas autarquias, foi aplicada toda a solução proposta para o OSIRIS-GOV. Em particular, foram recolhidos os atributos demográficos necessários, foi mapeada a cadeia de fornecimento publicamente observável e foram identificadas as fugas de dados com credenciais expostas de cada entidade, de modo a avaliar o ciber-risco de cada câmara municipal, descrito através da probabilidade de ocorrência de um ciberincidente.

A título de exemplo, a Figura 9 ilustra visualmente a cadeia de fornecimento mapeada pelo OSIRIS-GOV para uma das autarquias analisadas: a Câmara Municipal de Albufeira.

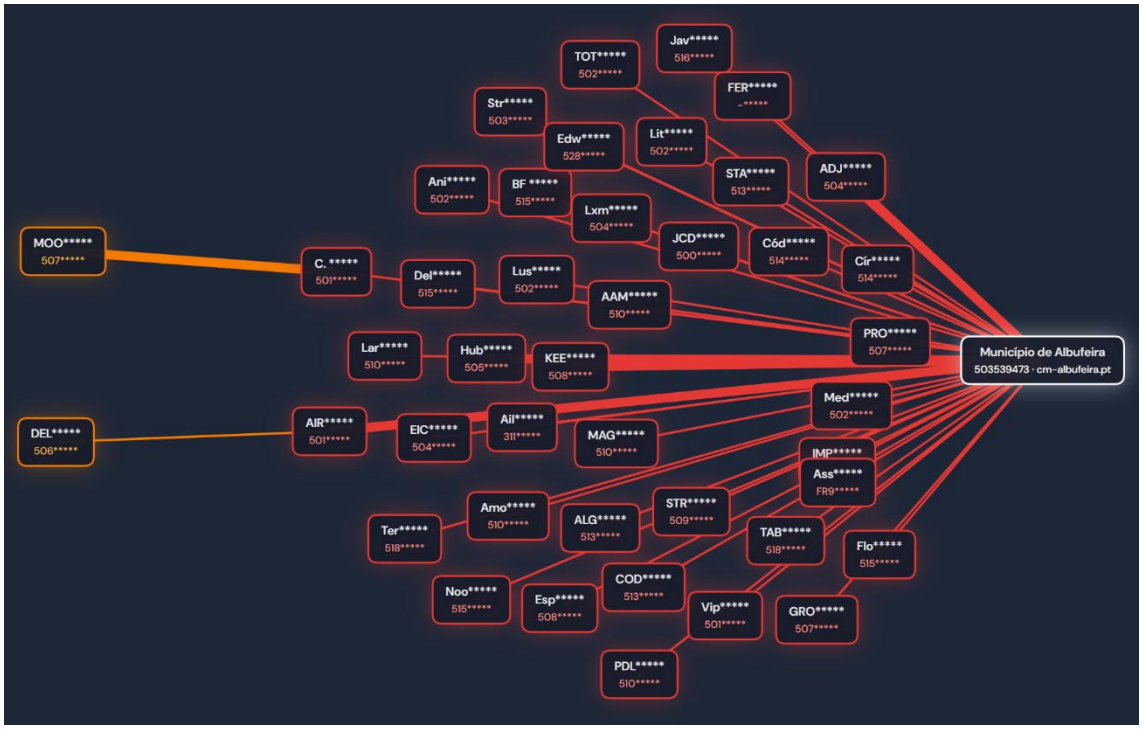

Figura 9 – Cadeia de Fornecimento da Câmara Municipal de Albufeira

Na Figura 9, apresenta-se o grafo determinado pelo OSIRIS-GOV com os terceiros, fornecedores e prestadores de serviços da Câmara Municipal de Albufeira.

Tal como exposto anteriormente, os vértices do grafo correspondem às entidades que fazem parte da cadeia de fornecimento da organização em análise, enquanto as arestas representam as relações contratuais existentes entre elas. As entidades integrantes da cadeia de fornecimento encontram-se anonimizadas.

Na imagem, as entidades do primeiro nível da cadeia de fornecimento, isto é, com relação contratual direta com a Câmara Municipal de Albufeira, encontram-se representadas a vermelho, enquanto as entidades do segundo nível da cadeia de fornecimento estão assinaladas a cor de laranja. A espessura de cada aresta representa o peso atribuído pelo OSIRIS-GOV à respetiva relação contratual, de acordo com o código CPV correspondente, conforme explicado na Tabela 15.

No caso de estudo realizado, de maneira a aferir o impacto da inclusão dos terceiros, fornecedores e prestadores de serviços na avaliação do ciber-risco para cada entidade, foi efetuada uma análise bipartida. Em primeiro lugar, cada câmara municipal foi submetida para análise enquanto entidade isolada, isto é, sem considerar a sua cadeia de fornecimento, mas apenas os atributos demográficos e as fugas de dados com credenciais expostas da própria autarquia. Em segundo lugar, foi incluída na análise a cadeia de fornecimento de cada autarquia, considerando os seus terceiros, fornecedores e prestadores de serviços, diretos e indiretos, de acordo com o proposto para o OSIRIS-GOV enquanto solução completa.

Com isto, procurou-se compreender o impacto da inclusão da cadeia de fornecimento publicamente observável na avaliação do ciber-risco organizacional, sendo este um dos fatores diferenciadores da solução proposta.

5.2. Resultados

De acordo com o enquadramento anterior, foi aplicada a solução OSIRIS-GOV a 24 câmaras municipais, para avaliação do ciber-risco associado.

Sendo o resultado da aplicação do OSIRIS-GOV a probabilidade de ocorrência de um ciberincidente num dado intervalo de tempo futuro, definiu-se como objetivo obter a probabilidade de ocorrência de um ciberincidente no ano de 2026. Para isso, a solução desenvolvida recolheu os dados correspondentes aos cinco anos anteriores a 2026, ou seja, desde 2021 até 2025, inclusive.

Os resultados obtidos encontram-se na Tabela 23. A coluna “Entidade Isolada” corresponde à análise de cada autarquia enquanto entidade isolada, isto é, sem considerar a respetiva cadeia de fornecimento, enquanto a coluna “Entidade e Cadeia de Fornecimento” mostra o valor retornado pelo OSIRIS-GOV após a inclusão da cadeia de fornecimento de cada entidade. Cada coluna contém a probabilidade de ocorrência de um ciberincidente em 2026, resultante da aplicação do OSIRIS-GOV.

Câmara Municipal	Entidade Isolada	Entidade e Cadeia de Fornecimento	Diferença
Albufeira	38%	46%	+8 p.p.
Almada	40%	46%	+6 p.p.
Aveiro	0%	16%	+16 p.p.
Borba	0%	10%	+10 p.p.
Bragança	6%	16%	+10 p.p.
Caminha	4%	16%	+12 p.p.
Castelo de Paiva	0%	16%	+16 p.p.
Cinfães	4%	8%	+4 p.p.
Faro	38%	46%	+8 p.p.
Leiria	38%	46%	+8 p.p.
Lisboa	42%	46%	+4 p.p.
Loures	38%	46%	+8 p.p.
Mirandela	0%	16%	+16 p.p.
Paços de Ferreira	2%	16%	+14 p.p.
Ponta Delgada	4%	16%	+12 p.p.
Porto	46%	46%	+0 p.p.
Santarém	2%	16%	+14 p.p.
Santo Tirso	2%	14%	+12 p.p.
Seixal	44%	46%	+2 p.p.
Sintra	40%	46%	+6 p.p.
Viana do Castelo	44%	46%	+2 p.p.
Vila do Conde	34%	46%	+12 p.p.
Vila Nova de Gaia	38%	46%	+8 p.p.
Vizela	2%	16%	+14 p.p.
Média	21%	30%	+9 p.p.

Tabela 23 – Resultados do Caso de Estudo

Conforme evidenciam os resultados expostos na Tabela 23, a inclusão dos terceiros, fornecedores e prestadores de serviços na avaliação do ciber-risco de cada organização tem tendência a aumentar a probabilidade de ocorrência de um ciberincidente, de acordo com a avaliação efetuada pelo OSIRIS-GOV.

Efetivamente, os resultados mostram que, em 23 das 24 câmaras municipais analisadas neste caso de estudo, a probabilidade de ocorrência de um ciberincidente determinada pelo OSIRIS-GOV – enquanto elemento descritor do ciber-risco avaliado – aumentou com a inclusão da cadeia de fornecimento. Em média, este valor situa-se nos 21% quando cada entidade é considerada de forma isolada, subindo nove pontos percentuais, para os 30%, quando também é considerada a respetiva cadeia de fornecimento. Deste modo, os resultados sugerem que o impacto dos terceiros, fornecedores e prestadores de serviços no ciber-risco avaliado para cada entidade não é negligenciável, pelo que deve ser considerado pelas soluções que se propõem a aferir o ciber-risco organizacional, como é o caso do OSIRIS-GOV.

Além desta análise agregada global, é possível distinguir os resultados obtidos pelas câmaras municipais consideradas de grande dimensão (com mais de mil colaboradores), em comparação com as câmaras municipais de pequena dimensão (com menos de mil colaboradores).

A Tabela 24 sintetiza estes resultados, apresentando os valores médios obtidos para as câmaras municipais, discriminadas pela respetiva dimensão.

Câmaras Municipais	**Entidade Isolada (Média)**	**Entidade e Cadeia de Fornecimento (Média)**	**Diferença (Média)**
Grandes (mais de mil colaboradores)	40%	46%	+6 p.p.
Pequenas (menos de mil colaboradores)	2%	15%	+13 p.p.

Tabela 24 – Resultados do Caso de Estudo por Dimensão das Câmaras Municipais

Tal como é visível na Tabela 24, existe uma diferença considerável entre os resultados obtidos para as autarquias com mais de mil colaboradores e as autarquias com menos de mil colaboradores.

Nas primeiras, a média dos valores obtidos para a probabilidade estimada de ocorrência de um ciberincidente é de 40% quando a autarquia é avaliada isoladamente e de 46% quando a cadeia de fornecimento também é tida em consideração, num aumento de seis pontos percentuais. Nas segundas, o ciber-risco avaliado para cada entidade isolada é descrito pela probabilidade de 2% de ocorrência de um ciberincidente em 2026, mas sobe para 15% com a inclusão da respetiva cadeia de fornecimento, aumentando 13 pontos percentuais. Estes resultados permitem extrair duas conclusões.

Em primeiro lugar, é claramente visível que as câmaras municipais com mais de mil colaboradores estão, à luz do OSIRIS-GOV, sujeitas a um maior risco de ocorrência de um ciberincidente em 2026, em comparação com as câmaras municipais com menos de mil colaboradores. Este comportamento poderá dever-se ao facto de as autarquias de maior dimensão terem associadas uma maior exposição de credenciais em fugas de dados, em virtude do maior número de colaboradores.

Efetivamente, esta observação encontra respaldo nas conclusões do mais recente Relatório Cibersegurança em Portugal, tema Sociedade [73], desenvolvido pelo Observatório de Cibersegurança do Centro Nacional de Cibersegurança (CNCS). Conforme evidencia este documento, a percentagem de organizações que sofreram ciberincidentes tem tendência a aumentar com o aumento da respetiva dimensão. De acordo com os dados observados, cerca de 12% das organizações portuguesas com 10 a 49 colaboradores sofreram um ciberincidente em 2024, enquanto, no mesmo período, este valor é de 17% para as organizações com 50 a 249 colaboradores e 20% para as organizações com 250 ou mais colaboradores, o que suporta a conclusão de que a suscetibilidade a ciberincidentes tende a aumentar com o número de colaboradores.

Em segundo lugar, os resultados também evidenciam uma diferença relevante no impacto da inclusão da cadeia de fornecimento, com uma preponderância maior da mesma no caso das autarquias de menor dimensão, num aumento superior ao dobro do observado no caso das autarquias de maior dimensão, em termos de pontos percentuais. Esta diferença sugere que, no caso das autarquias menores, o principal risco reside nos seus terceiros, fornecedores e prestadores de serviços, possivelmente devido à maior dependência destas entidades, enquanto, no caso das autarquias maiores, o ciber-risco está, essencialmente, concentrado na própria autarquia.

Em suma, os resultados obtidos pelo presente caso de estudo no contexto da administração pública local portuguesa permitiram, por um lado, demonstrar a viabilidade da aplicação prática do OSIRIS-GOV e, por outro lado, ilustrar a importância de considerar não só cada entidade individualmente, mas também os seus terceiros, fornecedores e prestadores de serviços na avaliação do ciber-risco organizacional.

6. Conclusão

Este capítulo conclui o trabalho realizado nesta dissertação, resumindo todo o processo seguido e destacando as principais ilações dele retiradas, bem como elencando alguns desafios e limitações que levantam possibilidades de eventual trabalho futuro.

6.1. Resumo e Principais Conclusões

Esta dissertação abordou a avaliação do ciber-risco organizacional tendo por base o recurso exclusivo a informação pública e a inclusão do risco de terceiros, fornecedores e prestadores de serviços nesta avaliação. Para esse efeito, foram seguidos os três objetivos delineados inicialmente, tendo sido todos cumpridos com sucesso.

Em primeiro lugar, indo ao encontro do proposto no Objetivo 1, foi realizada uma revisão de literatura – seguindo a metodologia PRISMA – para levantamento do estado da arte no que concerne à aplicação de um processo de OSINT para recolher e tratar informação útil de modo a avaliar o ciber-risco organizacional, considerando o risco de terceiros e o seu potencial de propagação ao longo da cadeia de fornecimento. Por um lado, esta revisão de literatura contemplou as soluções atualmente existentes para compreender o ciber-risco com base em informação pública, detalhando o processo de OSINT que lhes está subjacente, desde a identificação das fontes de informação até à apresentação de resultados, passando pela recolha, pelo tratamento e pela análise de dados. Por outro lado, foram também estudados os pilares essenciais da avaliação do risco de terceiros e da propagação do ciber-risco ao longo da cadeia de fornecimento, destacando as abordagens propostas na literatura para esta modelação. Em ambos os casos, foram elencadas as principais falhas, lacunas e limitações identificadas. Em particular, destacou-se a tendência para dissociar o risco dos terceiros, fornecedores e prestadores de serviços do ciber-risco da própria organização, bem como a falta de visibilidade sobre a cadeia de fornecimento, raramente indo para além de um primeiro nível de relações contratuais diretas. A par disto, acresceu ainda a preferência da maior parte das soluções estudadas pela identificação de vulnerabilidades conhecidas nos ativos expostos à *Internet* enquanto principal fator aferidor do ciber-risco, deixando em aberto a exploração de outros vetores de risco mais alinhados com as tendências dos ciberataques atuais.

Em segundo lugar, de acordo com o Objetivo 2, desenvolveu-se uma ferramenta – denominada OSIRIS-GOV – capaz de avaliar o ciber-risco organizacional, tendo em consideração a cadeia de fornecimento publicamente observável das entidades em análise. O desenvolvimento desta solução partiu das falhas, lacunas e limitações identificadas anteriormente, tendo culminado na definição de uma arquitetura e na subsequente implementação. A ferramenta desenvolvida considerou a exposição de credenciais corporativas em fugas de dados como elemento aferidor do ciber-risco organizacional, bem como a informação demográfica sobre as organizações. Estas decisões partiram, respetivamente, das evidências apresentadas em relatórios de ciberincidentes e da análise de algumas soluções já existentes, identificadas na revisão de literatura para levantamento do estado da arte. A solução proposta visou, através de um classificador inteligente, prever a probabilidade de ocorrência de um ciberincidente numa organização, com base em informação pública relativa à própria organização e à respetiva cadeia de fornecimento. Os resultados obtidos evidenciaram uma melhoria na capacidade de previsão de ocorrência de um ciberincidente, em comparação com abordagens semelhantes às estudadas na literatura recolhida.

Por último, a solução desenvolvida foi testada e validada num contexto real, de acordo com o Objetivo 3. Para esse efeito, foi efetuado um caso de estudo da aplicação do OSIRIS-GOV à administração pública local portuguesa. Neste caso de estudo, foram avaliados os resultados obtidos para 24 câmaras municipais, que evidenciaram a importância de incluir os terceiros, fornecedores e prestadores de serviços na avaliação do ciber-risco organizacional, em vez de considerar cada organização como uma entidade isolada. Esta necessidade de inclusão da cadeia de fornecimento na análise mostrou-se particularmente relevante para as autarquias de menor dimensão.

Em conclusão, esta dissertação cumpriu os três objetivos propostos, ao conseguir avaliar o ciber-risco de uma organização com base em informação pública, tendo em conta os seus terceiros, fornecedores e prestadores de serviços, diretos ou indiretos.

6.2. Desafios, Limitações e Trabalho Futuro

Apesar de os objetivos delineados terem sido cumpridos pelo trabalho realizado, surgiram alguns desafios e limitações que suscitam a possibilidade de continuação futura do trabalho desenvolvido.

Em primeiro lugar, a utilização da VCDB enquanto única origem do subconjunto de dados de organizações vítimas de ciberincidentes acarreta uma limitação. Tendo em conta que esta base de dados representa as organizações que foram vítimas de um ciberincidente tornado público, mas que existirão muitas organizações vítimas de ciberincidentes que não se tornaram públicos, isto pode introduzir um enviesamento nos dados fornecidos ao classificador desenvolvido. Por isso, como potencial resolução para este problema, abre-se a possibilidade de recorrer a outras fontes de informação como forma de identificação de organizações que sofreram ciberincidentes, de modo a melhorar a solução proposta.

Em segundo lugar, embora a opção pela exposição de credenciais corporativas em fugas de dados – em vez das vulnerabilidades detetadas em ativos expostos à *Internet* – enquanto fator aferidor do ciber-risco organizacional seja devidamente fundamentada e validada, ambos os indicadores poderiam ser considerados em simultâneo. Nesta medida, numa eventual extensão do trabalho desenvolvido, a informação atualmente recolhida poderia ser complementada por esse ou outro indicador do ciber-risco, com o intuito de vir a obter melhores resultados.

Em terceiro lugar, a aplicabilidade da solução encontra-se – por força dos requisitos delineados e do contexto em que se insere – limitada a organizações cuja cadeia de fornecimento seja passível de ser observada publicamente, pelas fontes de informação selecionadas. No entanto, existem muitas organizações que beneficiariam da aplicação de uma solução como o OSIRIS-GOV, mas que não têm os seus terceiros, fornecedores e prestadores de serviços consultáveis nas fontes de informação utilizadas. Como tal, a solução atual poderia ser estendida para utilizar outros meios de recolha da informação necessária para a determinação da cadeia de fornecimento, de maneira a aumentar a sua abrangência e aplicabilidade prática.

Por fim, a descrição do ciber-risco como a probabilidade de ocorrência de um ciberincidente num intervalo de tempo futuro, apesar de ser um indicador útil e concreto, por ser independente da avaliação subjetiva por parte de especialistas, poderia ser substituída ou complementada por outro indicador, com igual ou melhor capacidade de tradução do ciber-risco. Por exemplo, pode vir a ser seguida uma abordagem de quantificação do ciber-risco em termos financeiros a partir desta probabilidade, sendo este um indicador potencialmente mais adequado para comunicar com os órgãos de gestão, responsáveis pela tomada de decisões que visam garantir a cibersegurança das organizações.

Em resumo, ainda que o trabalho proposto tenha sido concluído com sucesso, subsistem linhas de desenvolvimento futuro que poderiam complementar e melhorar a solução OSIRIS-GOV.

Referências Bibliográficas

[1] European Commission, «Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions - State of the Digital Decade 2024». 2 de julho de 2024. Acedido: 6 de outubro de 2025. [Em linha]. Disponível em: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52024DC0260

[2] OECD, «OECD Digital Economy Outlook 2024 (Volume 2): Strengthening Connectivity, Innovation and Trust», OECD Publishing, nov. 2024. doi: 10.1787/3adf705b-en.

[3] European Union Agency for Cybersecurity (ENISA), «ENISA Threat Landscape 2024», Publications Office of the European Union, LU, set. 2024. doi: 10.2824/0710888.

[4] European Union Agency for Cybersecurity (ENISA), «ENISA Threat Landscape 2025», Publications Office of the European Union, LU, out. 2025. doi: 10.2824/1946374.

[5] Check Point Research, «Global Cyber Attacks Surge 21% in Q2 2025 — Europe Experiences the Highest Increase of All Regions», Check Point Blog. Acedido: 5 de outubro de 2025. [Em linha]. Disponível em: https://blog.checkpoint.com/research/global-cyber-attacks-surge-21-in-q2-2025-europe-experiences-the-highest-increase-of-all-regions/

[6] J. Ruohonen, J. L. Nielsen, e J. Skórczynski, «Risks and Compliance with the EU's Core Cyber Security Legislation», 29 de agosto de 2025, *arXiv*: arXiv:2508.21386. doi: 10.48550/arXiv.2508.21386.

[7] European Union Agency for Cybersecurity (ENISA), «NIS Investments 2025», Publications Office of the European Union, LU, dez. 2025. doi: 10.2824/7442427.

[8] A. Folorunso, I. Wada, B. Samuel, e V. Mohammed, «Security compliance and its implication for cybersecurity», *World J. Adv. Res. Rev.*, vol. 24, n.º 1, pp. 2105–2121, out. 2024, doi: 10.30574/wjarr.2024.24.1.3170.

[9] International Organization for Standardization (ISO) e International Electrotechnical Commission (IEC), *ISO/IEC 27001:2022 Information security, cybersecurity and privacy protection — Information security management systems — Requirements*, outubro de 2022. Acedido: 5 de outubro de 2025. [Em linha]. Disponível em: https://www.iso.org/standard/27001

[10] Global Standards, «ISO Survey 2024», Global Standards, 2025. Acedido: 23 de janeiro de 2026. [Em linha]. Disponível em: https://www.globalstd.com/wp-content/uploads/2025/10/Infografia-ISO-Survey-2024_EN.pdf

[11] Instituto Português de Acreditação, I.P. (IPAC), «Base de Dados Nacional - Sistemas de Gestão Certificados», IPAC. Acedido: 28 de fevereiro de 2026. [Em linha]. Disponível em: http://www.ipac.pt/pesquisa/pesq_empcertif.asp

[12] Parlamento Europeu e Conselho da União Europeia, *Diretiva (UE) 2022/2555 do Parlamento Europeu e do Conselho*. 2022. Acedido: 5 de outubro de 2025. [Em linha]. Disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:32022L2555

[13] República Portuguesa, *Decreto-Lei n.º 125/2025, de 4 de dezembro*. 2025. Acedido: 23 de janeiro de 2026. [Em linha]. Disponível em: https://files.diariodarepublica.pt/1s/2025/12/23400/0000400068.pdf

[14] Parlamento Europeu e Conselho da União Europeia, *Regulamento (UE) 2022/2554 do Parlamento Europeu e do Conselho*. 2022. Acedido: 5 de outubro de 2025. [Em linha]. Disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:32022R2554

[15] European Union Agency for Cybersecurity (ENISA), «ENISA Threat Landscape for Supply Chain Attacks», Publications Office of the European Union, LU, jul. 2021. doi: 10.2824/168593.

[16] A. Yadav, A. Kumar, e V. Singh, «Open-source intelligence: a comprehensive review of the current state, applications and future perspectives in cyber security», *Artif Intell Rev*, vol. 56, n.º 11, pp. 12407–12438, nov. 2023, doi: 10.1007/s10462-023-10454-y.

[17] A. Qusef e H. Alkilani, «The effect of ISO/IEC 27001 standard over open-source intelligence», *PeerJ Computer Science*, vol. 8, p. e810, jan. 2022, doi: 10.7717/peerj-cs.810.

[18] H. AlKilani e A. Qusef, «OSINT Techniques Integration with Risk Assessment ISO/IEC 27001», em *International Conference on Data Science, E-learning and Information Systems 2021*, Ma'an Jordan: ACM, abr. 2021, pp. 82–86. doi: 10.1145/3460620.3460736.

[19] D. Govardhan, G. G. S. H. Krishna, V. Charan, S. V. A. Sai, e R. R. Chintala, «Key Challenges and Limitations of the OSINT Framework in the Context of Cybersecurity», em *2023 2nd International Conference on Edge Computing and Applications (ICECAA)*, Namakkal, India: IEEE, jul. 2023, pp. 236–243. doi: 10.1109/ICECAA58104.2023.10212168.

[20] M. J. Page *et al.*, «The PRISMA 2020 statement: an updated guideline for reporting systematic reviews», *BMJ*, p. n71, mar. 2021, doi: 10.1136/bmj.n71.

[21] M. J. Page *et al.*, «PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews», *BMJ*, p. n160, mar. 2021, doi: 10.1136/bmj.n160.

[22] A. Bahji, L. Acion, A.-M. Laslett, e B. Adinoff, «Exclusion of the non-English-speaking world from the scientific literature: Recommendations for change for addiction journals and publishers», *Nordic Studies on Alcohol and Drugs*, vol. 40, n.º 1, pp. 6–13, fev. 2023, doi: 10.1177/14550725221102227.

[23] K. Stouffer, T. Zimmerman, C. Tang, J. Lubell, J. Cichonski, e J. McCarthy, «Cybersecurity framework manufacturing profile», National Institute of Standards and Technology, Gaithersburg, MD, NIST IR 8183, set. 2017. doi: 10.6028/NIST.IR.8183.

[24] R. Ross, V. Pillitteri, R. Graubart, D. Bodeau, e R. McQuaid, «Developing cyber-resilient systems: a systems security engineering approach», National Institute of Standards and Technology, Gaithersburg, MD, NIST SP 800-160v2r1, dez. 2021. doi: 10.6028/NIST.SP.800-160v2r1.

[25] Institute of Risk Management, *Cyber Risk Resources for Practitioners*. Institute of Risk Management, 2014. Acedido: 27 de dezembro de 2025. [Em linha]. Disponível em: https://www.theirm.org/media/7237/irm-cyber-risk-resources-for-practitioners.pdf

[26] Y.-W. Hwang, I.-Y. Lee, H. Kim, H. Lee, e D. Kim, «Current Status and Security Trend of OSINT», *Wireless Communications and Mobile Computing*, vol. 2022, n.º 1, p. 1290129, jan. 2022, doi: 10.1155/2022/1290129.

[27] Z. Avrahami, M. Zwilling, e C. Hajaj, «Leveraging OSINT for Advanced Proactive Cybersecurity: Strategies and Solutions», *IEEE Access*, vol. 13, pp. 154229–154250, 2025, doi: 10.1109/ACCESS.2025.3603868.

[28] J. Pastor-Galindo, P. Nespoli, F. Gomez Marmol, e G. Martinez Perez, «The Not Yet Exploited Goldmine of OSINT: Opportunities, Open Challenges and Future Trends», *IEEE Access*, vol. 8, pp. 10282–10304, 2020, doi: 10.1109/ACCESS.2020.2965257.

[29] C. Sauerwein, I. Pekaric, M. Felderer, e R. Breu, «An analysis and classification of public information security data sources used in research and practice», *Computers & Security*, vol. 82, pp. 140–155, mai. 2019, doi: 10.1016/j.cose.2018.12.011.

[30] T. Babenko, K. Kolesnikova, O. Abramkina, e Y. Vitulyova, «Automated OSINT Techniques for Digital Asset Discovery and Cyber Risk Assessment», *Computers*, vol. 14, n.º 10, p. 430, out. 2025, doi: 10.3390/computers14100430.

[31] A. Daskevics e A. Nikiforova, «ShoBeVODSDT: Shodan and Binary Edge based vulnerable open data sources detection tool or what Internet of Things Search Engines know about you», em *2021 Second International Conference on Intelligent Data Science Technologies and Applications (IDSTA)*, Tartu, Estonia: IEEE, nov. 2021, pp. 38–45. doi: 10.1109/IDSTA53674.2021.9660818.

[32] S. Silvestri, G. Tricomi, G. F. Russo, e M. Ciampi, «A Natural Language Processing-based Approach for Cyber Risk Assessment in the Healthcare Ecosystems», em *Ital-IA 2024: 4th National Conference on Artificial Intelligence*, Naples, Italy: CEUR Workshop Proceedings, mai. 2024, pp. 141–146. Acedido: 27 de dezembro de 2025. [Em linha]. Disponível em: https://ceur-ws.org/Vol-3762/527.pdf

[33] J.-Z. Yang, F. Liu, Y.-J. Zhao, L.-L. Liang, e J.-Y. Qi, «NiNSRAPM: An Ensemble Learning Based Non-intrusive Network Security Risk Assessment Prediction Model», em *2022 7th IEEE International Conference on Data Science in Cyberspace (DSC)*, Guilin, China: IEEE, jul. 2022, pp. 17–23. doi: 10.1109/DSC55868.2022.00010.

[34] J. Yang, L. Liang, e J. Qi, «A Practical Non-Intrusive Cyber Security Vulnerability Assessment Method for Cyber-Insurance», em *2023 8th International Conference on Data Science in Cyberspace (DSC)*, Hefei, China: IEEE, ago. 2023, pp. 261–269. doi: 10.1109/DSC59305.2023.00045.

[35] D. Ribeiro, V. Fonte, L. F. Ramos, e J. M. Silva, «Assessing the information security posture of online public services worldwide: Technical insights, trends, and policy implications», *Government Information Quarterly*, vol. 42, n.º 2, p. 102031, jun. 2025, doi: 10.1016/j.giq.2025.102031.

[36] B. Genge, P. Haller, e C. Enachescu, «Beyond Internet Scanning: Banner Processing for Passive Software Vulnerability Assessment», *International Journal of Information Security Science*, vol. 4, n.º 3, pp. 81–91, set. 2015.

[37] Y. Liu *et al.*, «Cloudy with a Chance of Breach: Forecasting Cyber Security Incidents», em *24th USENIX Conference on Security Symposium*, Washington, D.C, USA: USENIX Association, ago. 2015, pp. 1009–1024. Acedido: 27 de dezembro de 2025. [Em linha]. Disponível em: https://dl.acm.org/doi/10.5555/2831143.2831207

[38] C. Harry, I. Sivan-Sevilla, e M. McDermott, «Measuring the size and severity of the integrated cyber attack surface across US county governments», *Journal of Cybersecurity*, vol. 11, n.º 1, p. tyae032, jan. 2025, doi: 10.1093/cybsec/tyae032.

[39] S. Lee e T. Shon, «Open source intelligence base cyber threat inspection framework for critical infrastructures», em *2016 Future Technologies Conference (FTC)*, San Francisco, CA, USA: IEEE, dez. 2016, pp. 1030–1033. doi: 10.1109/FTC.2016.7821730.

[40] A. Mukhopadhyay e K. Luther, «OSINT Clinic: Co-designing AI-Augmented Collaborative OSINT Investigations for Vulnerability Assessment», em *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, Yokohama Japan: ACM, abr. 2025, pp. 1–22. doi: 10.1145/3706598.3713283.

[41] D. R. Hayes e F. Cappa, «Open-source intelligence for risk assessment», *Business Horizons*, vol. 61, n.º 5, pp. 689–697, set. 2018, doi: 10.1016/j.bushor.2018.02.001.

[42] A. Sarabi, P. Naghizadeh, Y. Liu, e M. Liu, «Risky business: Fine-grained data breach prediction using business profiles», *J Cyber Secur*, vol. 2, n.º 1, pp. 15–28, dez. 2016, doi: 10.1093/cybsec/tyw004.

[43] D. Agostinho, N. Tulcidas, M. Aniakor, H. Miranda, P. M. Ferreira, e A. Respicio, «Poster: Quality Threat Intelligence to Enhance Cyber Security in the Public Administration», em *2025 23rd International Symposium on Network Computing and Applications (NCA)*, Lisbon, Portugal: IEEE, nov. 2025, pp. 326–327. doi: 10.1109/NCA67271.2025.00065.

[44] Verizon Business, «2025 Data Breach Investigations Report», Verizon Business, 2025. Acedido: 27 de dezembro de 2025. [Em linha]. Disponível em: https://www.verizon.com/business/resources/reports/2025-dbir-data-breach-investigations-report.pdf

[45] R. Ghioni, M. Taddeo, e L. Floridi, «Open source intelligence and AI: a systematic review of the GELSI literature», *AI & Soc*, vol. 39, n.º 4, pp. 1827–1842, ago. 2024, doi: 10.1007/s00146-023-01628-x.

[46] J. M. Boyens, «Cybersecurity Supply Chain Risk Management Practices for Systems and Organizations», National Institute of Standards and Technology, Gaithersburg, MD, NIST SP 800-161r1-upd1, 2024. doi: 10.6028/NIST.SP.800-161r1-upd1.

[47] Joint Task Force Interagency Working Group, «Security and Privacy Controls for Information Systems and Organizations», National Institute of Standards and Technology, set. 2020. doi: 10.6028/NIST.SP.800-53r5.

[48] C. Topping, A. Dwyer, O. Michalec, B. Craggs, e A. Rashid, «Beware suppliers bearing gifts!: Analysing coverage of supply chain cyber security in critical national infrastructure sectorial and cross-sectorial frameworks», *Computers & Security*, vol. 108, p. 102324, set. 2021, doi: 10.1016/j.cose.2021.102324.

[49] G. A. Weaver, M. Perrie, B. Rolston, D. Buddenbohm, e R. Erbes, «Obstacles to Practical Supply Chain Risk Management for Digital Components», *IEEE Secur. Privacy*, pp. 2–11, 2025, doi: 10.1109/MSEC.2025.3591358.

[50] S. Boyson, «Cyber supply chain risk management: Revolutionizing the strategic control of critical IT systems», *Technovation*, vol. 34, n.º 7, pp. 342–353, jul. 2014, doi: 10.1016/j.technovation.2014.02.001.

[51] A. Ghadge, M. Weiß, N. D. Caldwell, e R. Wilding, «Managing cyber risk in supply chains: A review and research agenda», *Supply Chain Management: An International Journal*, vol. 25, n.º 2, pp. 223–240, jul. 2019, doi: 10.1108/SCM-10-2018-0357.

[52] A. Dash, S. P. Sarmah, M. K. Tiwari, S. K. Jena, e C. H. Glock, «Cybersecurity investments in supply chains with two-stage risk propagation», *Computers & Industrial Engineering*, vol. 197, p. 110519, nov. 2024, doi: 10.1016/j.cie.2024.110519.

[53] Y. Li e L. Xu, «Cybersecurity investments in a two-echelon supply chain with third-party risk propagation», *International Journal of Production Research*, vol. 59, n.º 4, pp. 1216–1238, fev. 2021, doi: 10.1080/00207543.2020.1721591.

[54] E. Seid, D. Ilter, F. Blix, e O. Popov, «Cyber Supply Chain Resilience: Analyzing ISO, NIST, and NIS2 Frameworks for Mitigating Third-Party Risks», *Procedia Computer Science*, vol. 263, pp. 591–599, 2025, doi: 10.1016/j.procs.2025.07.071.

[55] J. Viega e J. B. Michael, «Struggling With Supply-Chain Security», *Computer*, vol. 54, n.º 7, pp. 98–104, jul. 2021, doi: 10.1109/MC.2021.3075412.

[56] A. Tzoneva, G. Momcheva, e B. Stoyanov, «Vendor Cybersecurity Risk Assessment in an Autonomous Mobility Ecosystem», em *2022 10th International Scientific Conference on Computer Science (COMSCI)*, Sofia, Bulgaria: IEEE, mai. 2022, pp. 1–7. doi: 10.1109/COMSCI55378.2022.9912588.

[57] H. Asılı e Ş. Bahtiyar, «SME-SHIELD: A Scalable AI Architecture for Third-Party Cyber Risk Assessment in Small and Medium Enterprises (SMEs)», em *2025 10th International Conference on Computer Science and Engineering (UBMK)*, Istanbul, Turkiye: IEEE, set. 2025, pp. 1374–1379. doi: 10.1109/UBMK67458.2025.11206972.

[58] B. S. Pinto, L. Cioffi, e F. Espósito, «Third-party Cloud Risk Management», em *2024 IEEE International Conference on Cyber Security and Resilience (CSR)*, London, United Kingdom: IEEE, set. 2024, pp. 445–451. doi: 10.1109/CSR61664.2024.10679395.

[59] O. F. Keskin, K. M. Caramancion, I. Tatar, O. Raza, e U. Tatar, «Cyber Third-Party Risk Management: A Comparison of Non-Intrusive Risk Scoring Reports», *Electronics*, vol. 10, n.º 10, p. 1168, mai. 2021, doi: 10.3390/electronics10101168.

[60] N. G. Filho, N. Rego, e J. Claro, «Supply chain flows and stocks as entry points for cyber-risks», *Procedia Computer Science*, vol. 181, pp. 261–268, 2021, doi: 10.1016/j.procs.2021.01.145.

[61] J. Deane, W. Baker, e L. Rees, «Cybersecurity in Supply Chains: Quantifying Risk», *Journal of Computer Information Systems*, vol. 63, n.º 3, pp. 507–521, mai. 2023, doi: 10.1080/08874417.2022.2081882.

[62] S. Mondal e R. Singh, «Cyber risk propagation and budget optimization in financial networks: a Monte Carlo approach», *Int J Syst Assur Eng Manag*, nov. 2025, doi: 10.1007/s13198-025-03050-6.

[63] G. Yakut e K. Uzan, «Strengthening Supply Chain Risk: A Proactive Prioritization Model», em *Proceedings of the Future Technologies Conference (FTC) 2024, Volume 2*, vol. 1155, Cham: Springer Nature Switzerland, 2024, pp. 87–98. doi: 10.1007/978-3-031-73122-8_6.

[64] P. Cardoso, «Quantifying Cyber Risk with FAIR - A Case-Study of Super Bock Bebidas, S.A.», Universidade Católica Portuguesa, 2025. Acedido: 1 de fevereiro de 2026. [Em linha]. Disponível em: http://hdl.handle.net/10400.14/56365

[65] Fundação Francisco Manuel dos Santos, «PORDATA», Sobre a Pordata | Pordata. Acedido: 9 de maio de 2026. [Em linha]. Disponível em: https://www.pordata.pt/sobre+a+pordata

[66] Intelligence X, «Intelligence X», intelx.io Help System. Acedido: 22 de fevereiro de 2026. [Em linha]. Disponível em: https://help.intelx.io/

[67] NIF.PT, «NIF.PT», NIF. Acedido: 30 de março de 2026. [Em linha]. Disponível em: https://www.nif.pt/

[68] Instituto dos Mercados Públicos, do Imobiliário e da Construção, I.P. (IMPIC), «O Portal Base», Base. Acedido: 22 de fevereiro de 2026. [Em linha]. Disponível em: https://www.base.gov.pt/Base4/pt/o-portal/base/

[69] chicoferreira, *chicoferreira/contratopublico*. (21 de fevereiro de 2026). Svelte. Acedido: 22 de fevereiro de 2026. [Em linha]. Disponível em: https://github.com/chicoferreira/contratopublico

[70] Comissão Europeia, *Regulamento (CE) n.º 213/2008 da Comissão*. 2007. Acedido: 29 de março de 2026. [Em linha]. Disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:32008R0213

[71] Verizon RISK Team, «The VERIS Community Database (VCDB)», The VERIS Community Database (VCDB). Acedido: 10 de maio de 2026. [Em linha]. Disponível em: https://verisframework.org/vcdb.html

[72] T. Hastie, R. Tibshirani, e J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2.ª ed. Springer, 2009.

[73] Centro Nacional de Cibersegurança (CNCS), «Relatório Cibersegurança em Portugal - Sociedade», Centro Nacional de Cibersegurança (CNCS), abr. 2026. Acedido: 6 de junho de 2026. [Em linha]. Disponível em: https://www.cncs.gov.pt/docs/rel-sociedade2025-obciber-cncs.pdf