Funded by
the European Union

FALCON is funded under the Horizon Europe
Framework Program
Grant Agreement ID 101121281

FALCON

Fighting Corruption
& Organised Crime

Deliverable D5.1

Title: Risk assessment, investigation, and decision support toolset (R1.0)

Dissemination Level:	PU - Public
Nature of the Deliverable:	R
Date:	31/10/2024
Work Package:	WP5 - Risk assessment, investigation and decision support tools
Editors:	ENG
Reviewers:	UPV, CENTRIC
Contributors:	ENG, ICCS, CERTH

Abstract: This document provides a preliminary overview of the tools being developed under Work Package 5 (WP5) of the FALCON project, aimed at enhancing anti-corruption efforts through innovative technologies. The deliverable outlines the development of three key tools: the Advanced Corruption Risk Assessment (ACRA) tool, the Predictive Analytics Tool, and the FALCON Dashboard and Pattern Analysis Tool. These tools are designed to support Law Enforcement Agencies (LEAs) and Anti-Corruption Authorities by offering real-time risk assessments, predictive insights into potential corruption cases, and decision-making support through pattern analysis. As the first version of the "Risk Assessment, Investigation, and Decision Support Toolset" this document lays the groundwork for future enhancements that will be detailed in subsequent project deliverables.

Disclaimer

Revision History

Date	Rev.	Description	Partner(s)
05/06/2024	0.1	ToC of the deliverable	ENG
14/06/2024	0.2	Section 1 completed	ENG
28/06/2024	0.3	Section 2 completed	ENG
02/07/2024	0.4	Introduction to section 3	ENG
04/07/2024	0.5	Integration of contribution from ICCS and CERTH	ENG, ICCS, CERTH
05/07/2024	0.6	First Draft of the deliverable	ENG
20/09/2024	0.7	Second Draft of the deliverable	ENG
16/10/2024	0.8	Sent to FALCON project reviewers	ENG
28/10/2024	0.9	Updated version based on reviewers' feedback	ENG, UPV, CENTRIC
31/10/2024	1.0	Final version ready for submission	ENG, ICCS

List of Authors

Partner	Author
ENG	Marco San Biagio, Andrea Barbetta
CERTH	Kostas Loumponias, George Koutidis
ICCS	Evgenia Adamopoulou, Nikolaos Peppes, Theodoros Alexakis, Emmanouil Daskalakis

Internal Reviewers

Partner	Reviewer(s)
UPV	Francisco Pérez
CENTRIC	Christopher Guiver

Table of Contents

Index of figures

Glossary

ACRA	Advanced Corruption Risk Assessment
AFA	French Anti-Corruption Agency
AI	Artificial Intelligence
ANAC	Anti-corruption National Authority (Autorità Nazionale Anticorruzione)
API	Application Programming Interface
ATECO	Economic Activity (Attività Economiche)
Bi-LSTM	Bidirectional Long Short-Term Memory
BDNCP	National Data base of Public Contracts
CIP	Corruption Intelligence Picture
CNFPT	National Center for Local Civil Service (Centre National de la Fonction Publique Territoriale)
CNN	Convolutional Neural Networks
COSP	Conference of the States Parties
CPI	Corruption Perceptions Index
DL	Deep Learning
DMS	Document Management System
EU	European Union
EUROPAM	European Public Accountability Mechanisms
GDELT	Global Database of Events, Language, and Tone
GUI	Graphical User Interface
ICT	Information and Communication Technologies
INSS	National Institute of Social Security
KB	Knowledge Base
KPI	Key Performance Indicator
LEA	Law Enforcement Agency
MET	European Tenders Monitoring
ML	Machine Learning
MOOC	Massive Open Online Course
NACAP	National Anti-Corruption Action Plan
NCPA	Network of Corruption Prevention Authorities

NLP	Natural Language Processing
NN	Neural Network
NTA	National Transparency Authority
RDF	Resource Description Framework
ReLU	Rectified Linear Unit
SALER	Rapid Alert System
SPARQL	Simple Protocol and RDF Query Language
SQL	Structured Query Language
SSH	Social Sciences and Humanities
TED	Tenders Electronic Daily
TRACK	Tools and Resources for Anti-Corruption Knowledge
UI	User Interface
UNCAC	United Nations Convention against Corruption
UNGASS	Special Sessions of the General Assembly against corruption
UNODC	United Nations Office on Drugs and Crime
WP	Work Package

Executive Summary

This document provides an overview of the tools being developed as part of Work Package 5 (WP5) within the FALCON project, which is focused on advancing anti-corruption efforts through innovative technologies. The deliverable details the scope and structure of the tools, their relevance to the broader FALCON project, and their potential impact on Law Enforcement Agencies (LEAs) and Anti-Corruption Authorities.

The document begins by establishing the importance of this deliverable within the FALCON project, connecting it to other Work Packages and highlighting its contribution to the overall goals of enhancing anti-corruption strategies. The scope of the project is contextualized within the European Union's broader fight against corruption, and this connection is further reinforced by an analysis of various initiatives and strategies employed across EU countries.

A detailed exploration of corruption in the EU is presented, including a definition of the term and an examination of the corruption index, with a focus on data related to EU countries. Additionally, the document discusses AI-based tools that governments are currently utilizing to combat corruption, showcasing the importance of leveraging advanced technology in this field.

The core of the document centres on the three key tools being developed within WP5. These tools are designed to provide real-time analysis, predictive insights, and decision-support capabilities tailored to the needs of LEAs and Anti-Corruption Authorities.

1. The **Advanced Corruption Risk Assessment (ACRA)** tool offers real-time analysis and comprehensive risk assessments, enabling authorities to identify and address corruption risks effectively.
2. The **Predictive Analytics Tool** focuses on forecasting potential instances of single bidding at the outset of tender processes, allowing for pre-emptive measures to mitigate corruption risks.
3. The **FALCON Dashboard and Pattern Analysis Tool** is designed to assist investigators by visualizing and analysing patterns within corruption cases, supporting informed decision-making throughout the investigative process.

These tools, developed under WP5, represent significant advancements in the fight against corruption and organized crime, leveraging cutting-edge technology to enhance the capabilities of anti-corruption authorities.

In summary, this document outlines the development of sophisticated tools under WP5 of the FALCON project, emphasizing their potential to significantly enhance the ability of LEAs and Anti-Corruption Authorities to combat corruption through real-time

analysis, risk assessment, and predictive analytics. This document represents Version 1 of the "Risk Assessment, Investigation, and Decision Support Toolset." Further details and advancements will be presented in the subsequent two documents that will be released as part of the ongoing project.

1. Introduction

1.1. Purpose of the Deliverable

This document constitutes the initial deliverable of Work Package 5 (WP5), titled “Risk Assessment, Investigation, and Decision Support Tools.” WP5 is dedicated to developing software tools designed to enhance various facets of corruption management. These tools will facilitate comprehensive risk assessment, support the investigation of corruption-related crimes, enable predictive analytics, and provide advanced visualization capabilities. Additionally, they will aid in decision-making processes and ensure the continuous and overall management of Corruption Intelligence Pictures (CIPs). The aim is to create an integrated suite of tools that empowers users to effectively monitor, analyse, and respond to corruption threats, thereby strengthening the integrity and transparency of organizational operations.

1.2. Relevance of D5.1 and connections with other Work Packages

Deliverable D5.1, titled "Risk Assessment, Investigation, and Decision Support Toolset (R1.0)," is strategically positioned within the FALCON project, offering an initial overview of the tools developed under this Work Package. As the first deliverable of WP5, it provides general information about the fight against corruption in the EU, examines existing tools available to support the work of law enforcement agencies (LEAs), and presents a preliminary overview of the new tools being developed. Subsequent deliverables, specifically "Risk Assessment, Investigation, and Decision Support Toolset (R2.0)" and "Risk Assessment, Investigation, and Decision Support Toolset (R3.0)," will offer more detailed information on the architecture and features of these tools.

WP5 is strictly connected to the following other Work Packages:

- WP5 is connected to WP3 for specifying the anti-corruption framework.
- WP5 is connected to WP3 because WP3 includes tasks focused on ensuring Trustworthy AI, a critical aspect for the development of AI tools in WP5. Additionally, WP3 is responsible for the installation and management of Keycloak, which supports secure user authentication and access control for the UI dashboard developed in WP5.
- WP5 (along with WP2) will collaborate with WP6 to build a strong community that will be informed and trained on corruption countermeasures.
- WP5 is closely linked to WP4, as it gathers data to be analysed in WP5.
- WP5 is also connected to WP7, like the other WPs, with a focus on communication and exploitation of results.

1.3. Structure of the deliverable

This deliverable is structured to present an initial overview of the tools developed in the context of Work Package 5 of the FALCON project. The document is structured in this way:

- **Section 1** presents an introduction to the scope of the deliverable and its structure. Its relevance within the broader project context, and its connection to other Work Packages and deliverables is underlined.
- **Section 2** shows some initiatives to fight corruption in EU countries. Starting from the definition of the term "corruption", the corruption index is presented analysing some data related to EU countries. Furthermore, some AI-based (Artificial Intelligence) tools used by governments against corruption are presented.
- **Section 3** introduces the tools being developed within the framework of FALCON WP5. Beginning with an overview of the WP5 architecture, which facilitates seamless interaction among its components, this section outlines the main tools. The ACRA (Advanced Corruption Risk Assessment) tool offers real-time analysis and comprehensive risk assessments specifically designed for Law Enforcement Agencies (LEAs) and Anti-Corruption Authorities. The predictive analytics tool for corruption cases aims to forecast potential instances of single bidding at the outset of tender processes. Lastly, the FALCON dashboard and pattern identification tool support investigators in making informed decisions regarding each corruption case.

2. Fight against Corruption

This chapter will introduce the definition of corruption and some initiatives to fight corruption in EU countries. This introduction is quite important to better understand the effectiveness of the tools developed in the context of WP5 to fight corruption.

2.1. Definition of Corruption

Corruption is a widespread issue that deeply affects societies, hindering social, economic, and political progress. It fosters social division, poverty, and inequality, while weakening democratic institutions and stifling economic growth. Additionally, it erodes public trust in institutions, perpetuating a cycle of incompetence and dissatisfaction. Corruption occurs when individuals or groups misuse their power for personal gain and can affect various sectors, such as business, politics, the judiciary, media, civil society, infrastructure, sports, health, and education. Examples include kickbacks in business, bribery in government, judicial bias, and media manipulation influenced by bribes.

Corruption generally involves a network of facilitators, such as bankers, attorneys, accountants, and real estate agents. By utilizing intricate financial tools and structures including opaque financial systems and anonymous shell corporations, these enablers aid in the hiding and laundering of criminal cash. This makes it challenging to track down the sources of unlawful income and gives dishonest people the freedom to enjoy their profits.

Understanding the dynamics of corruption can be facilitated through Klitgaard's formula (Klitgaard, 1988):

$$C = R + D - A$$

In this equation:

- **C** stands for corruption,
- **R** for economic rent,
- **D** for discretionary authority, and
- **A** for accountability.

This model posits that corruption increases with the opportunities for economic rent (R) and the amount of discretionary power (D) held by officials. Economic rent refers to the excess profits that can be earned in an environment where competition is restricted, creating a fertile ground for corrupt practices. Discretionary authority, on the other hand, pertains to the power officials must make decisions that can significantly benefit specific individuals or groups, often without sufficient oversight.

Conversely, corruption can be mitigated by increasing accountability (A). When officials are held responsible for their actions through robust oversight mechanisms, transparent processes, and effective enforcement of laws, the likelihood of engaging in corrupt practices diminishes. Accountability acts as a deterrent, reducing the appeal and feasibility of corrupt activities.

The formula also highlights the importance of systemic reforms in combating corruption. By reducing monopolies and carefully regulating discretion, transparency can be enhanced, thereby reducing opportunities for corruption. Additionally, increasing the probability of detection and the severity of penalties for corrupt behaviour can further dissuade individuals from engaging in such activities. Successful anti-corruption strategies often involve a combination of legal reforms, public awareness campaigns, and institutional strengthening to foster a culture of integrity and accountability.

Klitgaard's insights underscore that corruption is not merely a matter of personal ethics but is deeply embedded in the systems and structures within which individuals operate. Therefore, combating corruption requires a comprehensive approach that addresses the root causes and enablers of corrupt behaviour. This includes fostering a culture of transparency, enhancing the capacity of institutions to detect and punish corrupt acts, and engaging civil society in monitoring and advocating for accountability.

2.2. Corruption Perceptions Index

To better understand the concept of corruption and to explain how the tools developed in WP5 will assist law enforcement agencies (LEAs) in investigating potential corruption cases, it is essential to consider the *Corruption Perceptions Index* (CPI). This index, established by Transparency International, ranks countries based on expert assessments and surveys measuring perceived levels of public sector corruption. Central to the CPI is its definition of corruption as the "abuse of entrusted power for private gain," emphasizing the fundamental issue the index addresses. This introductory section provides context for the rationale behind developing these tools in WP5.

Scoring 180 countries around the world, the CPI is the leading global indicator of public sector corruption, scoring on a scale of 0 (highly corrupt) to 100 (very clean), as pointed out in Figure 1.

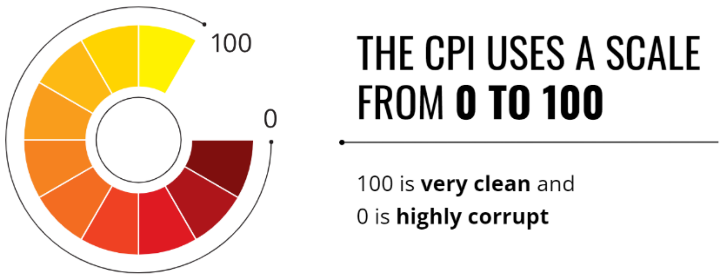

Figure 1 CPI scale: Values close to 0 indicate high levels of corruption, while values close to 100 indicate very low levels of corruption

Even though the CPI is now the most widely used indicator of corruption in the world, it is crucial to remember that it has certain limitations. First of all, the CPI does not distinguish between various kinds of corruption (some are not even included in the index); therefore, perceptions of corruption may not necessarily correspond to the genuine amount of corruption. In conjunction with further assessments, the CPI offers a more comprehensive view. Furthermore, the CPI is a superior tool for analysing long-term trends because perceptions tend to shift gradually.

The 2023 CPI shows that corruption is thriving across the world. Of the 180 countries that the CPI index measures, only 28 have seen a significant improvement in corruption over the past 12 years, while 34 have seen a marked decline. For 118 countries, there was no discernible change. Furthermore, corruption continues to be an issue for the majority of people worldwide because, according to Transparency International, over 80% of people live in nations whose CPI index is lower than the global average of 43.

Despite the fact that their ranks don't reflect it, high-ranking nations on the CPI struggle with impunity. Many cases of cross-border corruption have included companies from high-scoring countries that pay foreign officials in order to conduct business. Some have questioned experts who sell secrecy or assist dishonest foreign authorities in any manner. However, high-scoring countries often fail to prosecute both the transnational corrupt individuals and their accomplices.

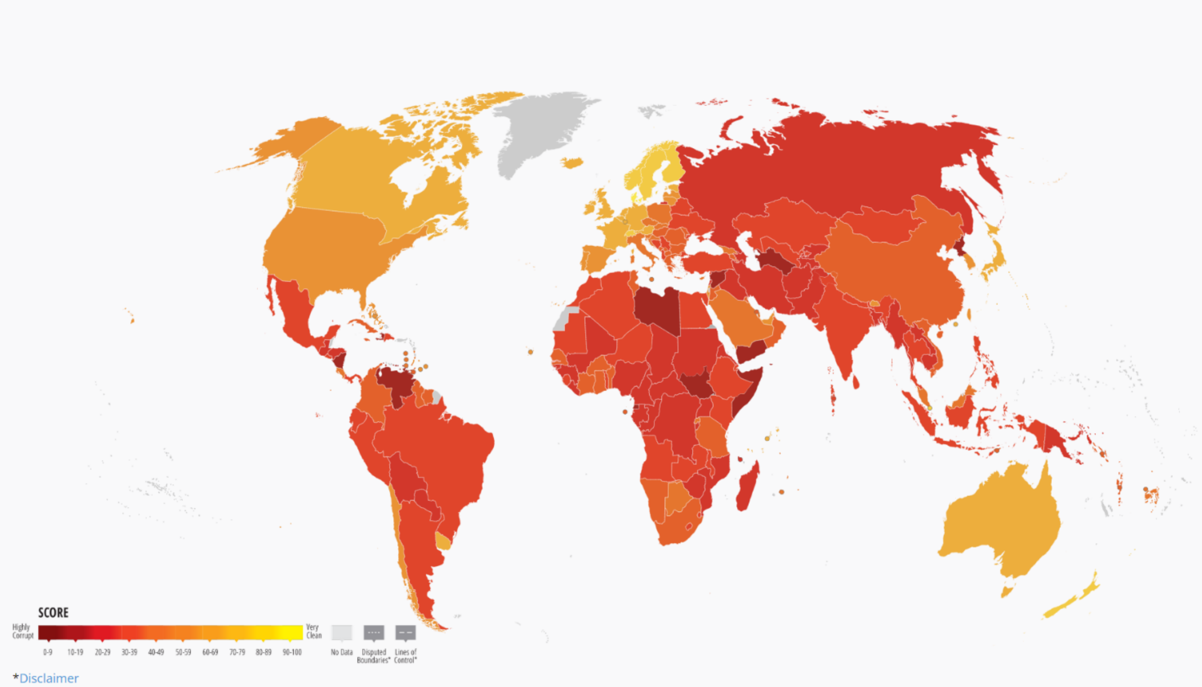

Figure 2 World map highlighting the CPI index per country. Perceptions of corruption are higher in regions of the world with darker colours

The figure indicates that perceptions of corruption are higher in regions of the world with darker colours. According to the CPI for 2023, it is determined that:

- The European Union and Western Europe continue to score highest overall, but regional average fell to 65 as political integrity and checks and balances deteriorated. Sub-Saharan Africa maintains the lowest average at 33, with democracy and the rule of law under threat, despite improvements in certain countries.
- All other regions of the world have averages below 50, indicating continued stagnation. The weak legal system, growing authoritarianism, and institutional corruption are issuing that Eastern Europe and Central Asia are facing.
- While some historically top countries are regressing, the Middle East and North Africa exhibit minimal improvement due to continuous struggles with political corruption and conflict, and the Asia Pacific region displays long-term stagnation. Ultimately, the Americas' pervasive impunity is being made possible by a lack of judicial independence and a feeble legal system.

Despite Western Europe consistently ranking high on the CPI, recent trends underscore why projects like FALCON, which develop tools to combat corruption and organized crime, are still crucial. The region's average score has dropped from 66 in 2021 to 65 in 2023, reflecting concerns over deteriorating political integrity and weakened checks and balances. Even high-scoring countries like the UK have seen significant declines, with the UK's score falling from 82 in 2017 to 71 in 2023 (Aa., Transparency International UK, 2024), driven by concerns about political corruption and public office abuse.

For these reasons, these trends suggest that while Europe may have relatively low levels of corruption compared to other regions, complacency or a decline in integrity could enable corrupt practices to take root. Initiatives like FALCON, which focus on advanced technologies for detecting corruption and organized crime, are essential to maintaining and strengthening governance, ensuring that corruption does not undermine public trust or erode democratic institutions.

2.3. Initiative to Combat Economic Crime and Corruption

Understanding the CPI provides a foundational perspective on global corruption levels. Building on this, various initiatives aim to combat economic crime and corruption at a more actionable level. One such initiative is led by the United Nations Office on Drugs and Crime (UNODC), which encourages and supports increased dialogue between businesses and governments on anti-corruption issues to establish transparency and sound partnerships.

UNODC has created a set of tools and services to furnish Member States, anti-corruption organizations, and the corporate sector with invaluable information regarding the UN Convention against Corruption (UNCAC). The first international anti-corruption treaty with global enforcement, the UNCAC, has measures that are applicable to enterprises everywhere.

The UNCAC is the most comprehensive and widely adopted anti-corruption accord that is legally binding on a global scale. However, while 190 countries have signed or ratified the convention, it is not universally adopted. Several nations, such as North Korea and Syria, have not signed the accord. Moreover, even among countries that have ratified it, implementation can be uneven. For example, Germany and Japan have taken longer to fully ratify key provisions related to asset recovery (Aa., Twenty years of UNCAC: uniting the world against corruption, s.d.). This underscores the importance of continued efforts to ensure broader and more consistent application of its principles worldwide.

The Conference of the States Parties (COSP), which assists States parties and signatories in implementing the Convention, is the main policy-making body of the Convention. It makes policy recommendations to UNODC for the creation and implementation of anti-corruption measures. A unique peer-review process called the Implementation Review Mechanism is used to evaluate how successfully States parties have actually adopted the Convention into their national legislation.

Companies can learn how to level the playing field and promote healthy business environments in the nations where they operate by using UNODC tools and resources.

UNODC provides also technical assistance in a number of corruption-related theme areas, including prevention, education, asset recovery, integrity in the criminal justice system, etc., to support States parties' efforts to fully implement the Convention.

The TRACK (Aa., TRACK — UNODC's central platform of tools and resources for anti-corruption knowledge, s.d.) portal is the central platform of “Tools and Resources for Anti-Corruption Knowledge” developed by UNODC. It features three primary characteristics:

1. *Legal Library:* This comprehensive database contains legislation and jurisprudence relevant to the United Nations Convention against Corruption (UNCAC) from over 175 countries. The information is organized by country, UNCAC article, legal system, and other criteria, making it a valuable resource for legal professionals and researchers.
2. *Anti-Corruption Learning Platform:* TRACK includes an educational component where users can access analytical tools and resources created by partner organizations. This platform supports continuous learning and knowledge-sharing among anti-corruption practitioners globally.
3. *Community of Practice: T*RACK fosters a collaborative environment where registered users, including anti-corruption experts and practitioners, can communicate, share information, and organize events. This feature enhances networking and the exchange of best practices within the anti-corruption community.

These features collectively support UNODC's mission to promote transparency, enhance legal knowledge, and facilitate international cooperation in the fight against corruption. To further these efforts, UNODC has developed a variety of tools, manuals, and publications designed to assist in combating corruption and economic crime. Continuously creating new resources, UNODC aims to improve understanding of the challenges, regulations, and best practices related to the implementation of UNCAC, as determined by the Review Mechanism. This revised platform, developed in accordance with the Convention's framework, provides valuable guidance to both the public and practitioners on navigating the legal complexities associated with anti-corruption efforts.

The TRACK portal's structure is designed to facilitate easy access to a wide range of resources. As illustrated in the figure, the portal is divided into three main sections: the Legal Library, the Bibliographic Database, and General Resources. The Legal Library includes resources categorized by chapters of the UNCAC, covering topics such as prevention, criminalization, law enforcement, international cooperation, and asset recovery. The Bibliographic Database organizes resources by thematic areas, including civil society, education and youth, environment, gender, international investment, private sector, public health, and sport. Additionally, the General Resources section provides overarching documents and guidelines related to the ratification and implementation of UNCAC, the Conference of the States Parties to UNCAC and its working groups, and Special Sessions of the General Assembly Against Corruption (UNGASS).

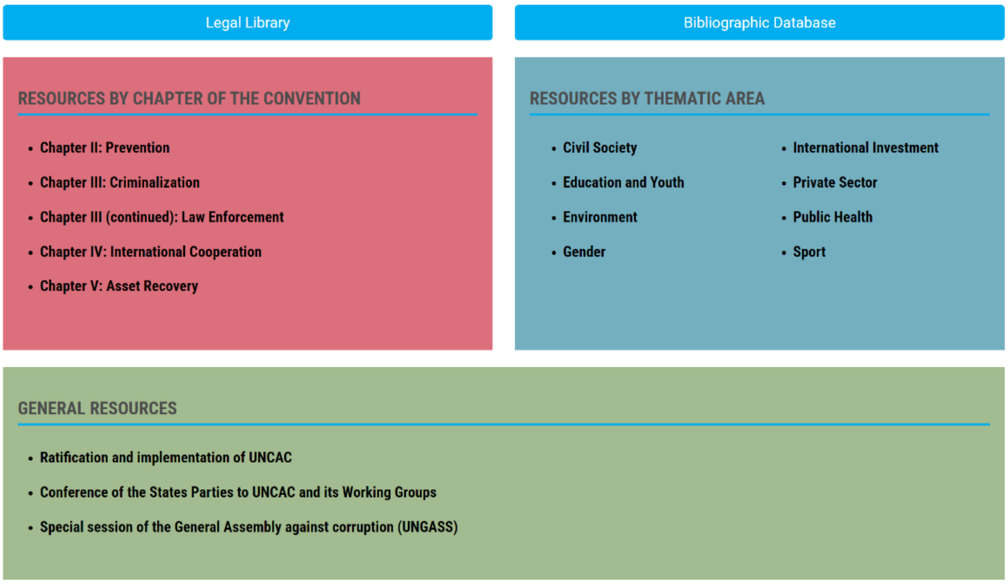

Figure 3 Overview of the TRACK portal's structure, showcasing the Legal Library, Bibliographic Database, and General Resources. The Legal Library categorizes resources by UNCAC chapters, while the Bibliographic Database organizes them by thematic areas, providing comprehensive support for anti-corruption efforts.

By organizing the portal in this manner, UNODC ensures that users can easily find the specific information they need, enhancing the overall effectiveness of their anti-corruption initiatives.

2.4. Artificial Intelligence and the fight against corruption

In the fight against corruption, artificial intelligence (AI) has emerged as a powerful technique in detecting criminal practices related to corruption itself. This technology provides tools to identify and stop fraudulent acts because of its ability to analyse huge amounts of data and uncover hidden patterns. While it has great potential, extendable to all countries, the actual application of AI is still quite limited and fraught with difficulties.

The integration of AI in anti-corruption strategies represents a significant advancement in the detection and prevention of corrupt practices. By leveraging AI, law enforcement agencies and anti-corruption bodies can enhance their capabilities to analyse complex datasets, predict potential corruption risks, and streamline their investigative processes. This modern approach complements the traditional initiatives led by organizations like UNODC, creating a multifaceted strategy in the global fight against corruption.

In the following paragraphs, we will explore the potential applications of artificial intelligence in combating corruption. We will highlight specific examples of AI-driven tools and technologies that have been developed to detect and prevent corrupt activities. By examining these innovative solutions, we aim to illustrate how AI can enhance the effectiveness of anti-corruption efforts and support law enforcement agencies in their mission to uphold integrity and transparency.

2.4.1. Potential uses of artificial intelligence

Corruption manifests in various forms and infiltrates numerous sectors, creating environments where unethical practices can thrive. To effectively combat this complex problem, it is essential to focus on the processes and procedures where corruption is most likely to occur. Technology, and specifically artificial intelligence (AI), offers significant promise in addressing these challenges by enhancing transparency, improving oversight, and streamlining investigative processes.

AI can analyse vast amounts of data to detect irregularities and patterns indicative of corrupt activities. By leveraging machine learning algorithms and advanced analytics, AI systems can identify suspicious transactions, monitor compliance, and provide real-time alerts to potential corruption risks. The integration of AI into anti-corruption strategies represents a transformative approach to preventing and mitigating corrupt practices.

The following are some potential uses of artificial intelligence:

- *Public Purchasing*: analysing public procurement processes to identify unusual patterns, such as customized tender specifications or repeated awards to the same business. Additionally, scrutinizing public contracts for irregular or potentially corrupt terms.

- *Public audits:* conduct budget audits to evaluate how municipalities and other public administrations allocate and spend their funds, identifying any unauthorized or irregular expenditures. Audit public grants and subsidies to detect mismatched or duplicated beneficiaries.

- *Finding patterns:* utilising machine learning to detect anomalous patterns by analysing large volumes of data, looking for indications of fraud. Text analysis can be employed to search for unusual terms or patterns in documents and emails, such as when investigating fraudulent reports.

- *Human Resources:* managing human resources to identify anomalies in internal promotions, transfers, and recruitment processes. Monitoring conflicts of interest by detecting potential links between public administration employees and private businesses.

- *Administration of public services:* Identifying inefficiencies, shortcomings, or unethical behaviour in the administration of public services. Ensuring the authenticity and consistency of documents submitted in administrative procedures, validating official documents, and preventing falsifications.

- *Database Integration:* Connecting public databases at local, national, or global levels to cross-reference information and enhance the detection of fraud through comprehensive data analysis.

- *Data Security:* Protecting sensitive information by alerting users to potential breaches or manipulation of official databases. Optimizing investigations by enabling AI to suggest courses of action based on prior cases when potential fraud is detected, thereby saving resources, and reducing investigation time.

- *Data structures:*
    - *Reporting systems:* AI-powered, anonymous platforms for reporting that ensure the safety of whistleblowers and the accuracy of the data they provide.
    - *Predictive systems:* Utilising previous data from the past and present trends, AI may be able to identify regions or industries that will likely have a higher fraud risk in the future. Public administration staff members and citizens should receive ethics and public integrity training that is customized to meet each person's unique needs as identified by AI.

- *Legislation and transparency*: AI tools that clarify the rationale behind business or government decisions can reduce the risk of undue favouritism and enhance transparency in decision-making processes.

- *Evaluation of laws and regulations:* AI is capable of comparing and analysing laws from various nations or areas in order to suggest modifications that close legal gaps that allow corruption.

In the next paragraphs, a set of AI-based tools to fight corruption are presented. Figure 4 below shows the list of tools described.

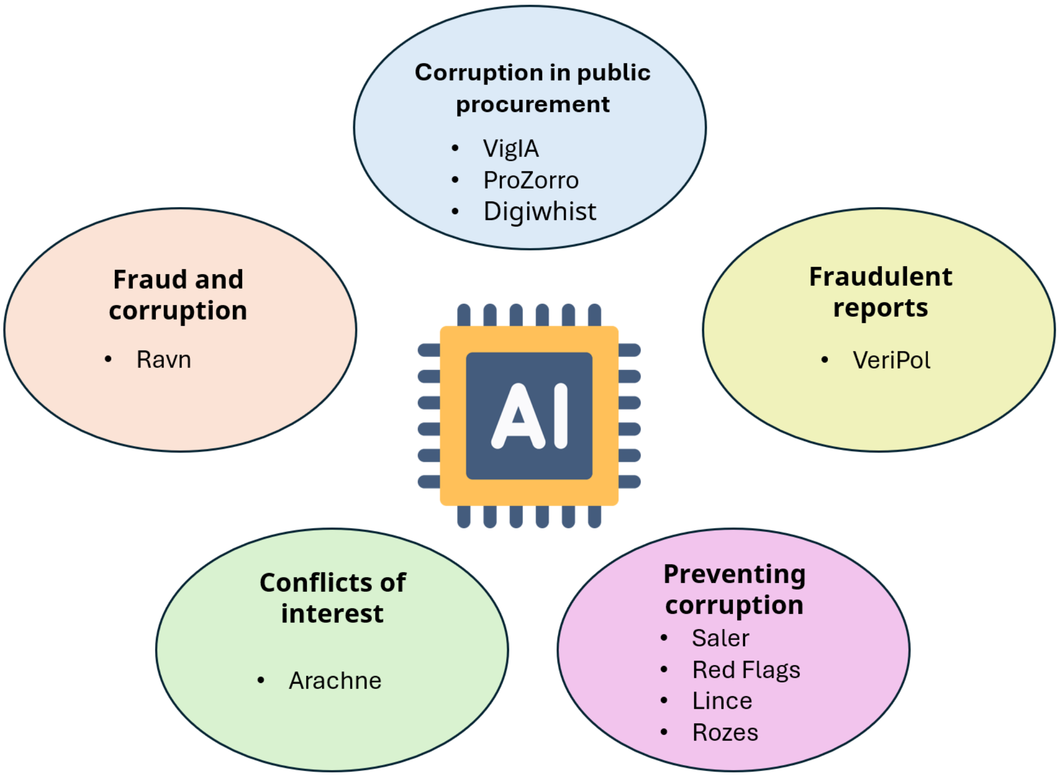

Figure 4 AI-based tools to fight corruption

2.4.2. Tools used against corruption

Having explored the potential uses of artificial intelligence in combating corruption, including its capabilities in explaining decisions and evaluating laws and regulations, the focus now shifts to specific tools that are being utilized in the fight against corruption. Some examples of artificial intelligence-based tools and applications currently used in the field of anti-fraud and anti-corruption are categorized below according to the purpose for which they are used. These tools, incorporating AI technology, demonstrate practical applications that enhance transparency, detect fraudulent activities, and support legal and regulatory reforms. This illustrates how advanced technologies are being harnessed to effectively address and mitigate corruption.

In the following, key examples of AI-based tools utilized in anti-corruption efforts are presented.

VigIA

The District Supervisory Office of Bogotá has integrated an innovative AI-powered tool named VigIA (Maas, s.d.), developed by Tic Tank of the University of Rosario (Argentina), into its anti-corruption strategy. This tool leverages data from the Electronic System for Public Procurement to pinpoint contracts within the Bogotá Mayor's Office that are susceptible to corruption and inefficiencies. Here's how VigIA functions and the benefits it offers:

- *Data Integration and Analysis:* VigIA utilizes advanced data analytics to systematically examine procurement data. By accessing information from the Electronic System for Public Procurement, it reviews various parameters of contracts to detect abnormalities or patterns that deviate from standard practices.
- *Risk Assessment:* The core of VigIA's functionality lies in its ability to assess the risk associated with each contract. Using machine learning algorithms, it categorizes contracts based on their vulnerability to corruption, flagging those that require closer scrutiny. This proactive approach helps in pre-empting corrupt activities by focusing oversight where it's most needed.
- *Efficiency Improvement:* Beyond identifying corruption vulnerabilities, VigIA also aims to enhance the efficiency of public spending. It analyzes the performance and outcomes of contracts to ensure that public funds are used effectively, thereby reducing wasteful expenditures, and improving project delivery.
- *Transparency and Accountability:* By making its findings accessible to relevant stakeholders, VigIA promotes transparency and accountability in public procurement. This transparency not only deters potential corrupt actions but also builds public trust in governmental operations.
- *Continuous Monitoring:* VigIA is designed for continuous monitoring of procurement processes. This ongoing oversight ensures that any irregularities are detected in a timely manner, allowing for immediate intervention.

ProZorro

The "ProZorro" system (Aa., Hertie school, s.d.), launched in Ukraine in 2016, represents a groundbreaking approach to enhancing transparency and combating corruption in government contracting through the use of technology. Developed through a collaborative effort involving international organizations, businesses, and civil society, ProZorro is an e-procurement system that harnesses artificial intelligence to scrutinize public procurement data, detect violations, and prevent the misuse of public funds. Key Features and Functions of ProZorro:

- *Transparency by Default:* ProZorro operates on a principle that all procurement data, except explicitly classified information, should be publicly accessible. This transparency ensures that anyone can see what is being purchased, by whom, from whom, and at what price.

- *Reverse Auction Mechanism:* The system uses a reverse auction format, which has been effective in reducing prices and increasing competition among suppliers. This mechanism allows multiple suppliers to bid prices down in real-time during the procurement process.
- *Hybrid Model:* ProZorro combines both centralized and decentralized elements. The system is centrally hosted but operates through various commercial marketplaces that connect to the central database. This model encourages competition among platforms while maintaining a unified database that ensures transparency and accessibility.
- *Open Source and Open API:* The technology behind ProZorro is open source, which allows for continuous improvement and adaptation by various developers. It also operates with an open API, enabling third-party developers to build applications that can interact with the ProZorro system.
- *Collaborative Development:* Developed in a unique collaboration among the government, business entities, and civil society organizations, ProZorro is an example of a multi-stakeholder approach to tackling corruption and improving efficiency in public procurement.
- *Impact and Recognition:* Since its implementation, ProZorro has saved the Ukrainian government and taxpayers significant amounts of money and has been recognized internationally, including receiving awards such as the World Procurement Award.

Digiwhist

Digiwhist (Aa., Digiwhist, s.d.) is an ambitious project funded by the European Union, designed to leverage big data in identifying fraud in public procurement across Europe. This initiative collaborates closely with anti-corruption organizations and generates both indicators and publicly accessible data to foster transparency and equity in public procurement. The tools developed by Digiwhist are available for free public use, promoting greater accountability and oversight. Key tools include:

- *OpenTender.eu:* A comprehensive platform that provides access to public procurement notices across 35 jurisdictions, enabling users to search and analyze procurement data easily.
- *European Public Accountability Mechanisms (EuroPAM):* This tool allows users to compare and contrast the regulations and laws governing public procurement in different European countries, fostering a better understanding of national compliance and accountability standards.

- *European Tenders Monitoring (MET)* (Aa., Digiwhist, s.d.)*:* Specialized software designed for risk assessment in public procurement procedures, helping to identify and mitigate potential risks associated with public contracts.

Ravn

Ravn (Aa., Smart Team Global, s.d.) is an advanced artificial intelligence software that surpasses human efficiency in filtering, indexing, and summarizing documents rapidly and accurately. Part of the burgeoning "Lawtech" sector, Ravn utilizes intelligent algorithms that learn from their experience, maximizing time and resources for legal professionals. These programs are capable of a range of tasks, such as automatically extracting passport numbers from photos or indexing international archives.

Despite their effectiveness, Ravn and similar AI tools are viewed not as a threat to employment but as valuable assets that increase the efficiency of the legal sector. By handling routine and time-consuming tasks, these AI solutions enable legal professionals to focus on more complex and strategic aspects of their work.

Some key features of Ravn are:

- *Increased Efficiency and Accuracy:* Ravn automates the processing of large volumes of documents, significantly reducing time and minimizing human error in tasks such as data extraction, indexing, and summarization.
- *Cost Savings and Scalability:* By automating routine tasks, Ravn reduces operational costs and can scale to handle extensive document archives, making it suitable for large organizations.
- *Enhanced Compliance and Risk Management:* Ravn helps ensure regulatory compliance by providing thorough analysis and documentation, flagging potential compliance issues, and assisting in effective risk management.

VeriPol

VeriPol (Aa., Universidad Complutense Madrid, s.d.) is an advanced AI system employed to identify fraudulent police reports. Developed in response to a surge in fictitious reports of violent robberies, VeriPol leverages machine learning and natural language processing techniques to analyse the language used in submitted complaints. The system's effectiveness has been demonstrated to be 91%, making it a powerful tool in distinguishing between genuine and fraudulent reports. By examining linguistic patterns and inconsistencies, VeriPol helps law enforcement agencies quickly and accurately identify deceitful claims, thereby improving resource allocation and enhancing public trust in the reporting process. Key features of VeriPol include:

- *Automated Linguistic Analysis:* VeriPol uses natural language processing (NLP) to analyse the linguistic features of police reports. This includes examining word choice, sentence structure, and patterns of speech that are indicative of deceit.

By analysing these elements, VeriPol can identify inconsistencies and signs of fabricated stories.

- *Behavioural Pattern Recognition:* By analysing historical data from both genuine and fraudulent reports, VeriPol's machine learning algorithms can recognize behavioural patterns commonly associated with false claims. This feature allows the system to detect anomalies and predict the likelihood of a report being fabricated based on past trends.
- *User-Friendly Interface for Law Enforcement:* VeriPol is designed with a practical interface that allows law enforcement officers to easily input reports and receive rapid assessments.

Arachne

Arachne (Aa., European Commission, s.d.) is a digital system created by the European Commission, to enhance oversight of projects funded by EU structural funds, including the European Social Fund and the European Regional Development Fund. Arachne detects possible fraud, conflict of interest, and irregularity risks by storing data from these projects and augmenting it with publicly available information. Some key benefits of Arachne are:

- *Fraud Detection and Prevention:* Arachne uses sophisticated algorithms to analyse data related to EU-funded projects, effectively identifying patterns that may indicate fraud, conflicts of interest, or other irregular activities. This helps in pre-emptively addressing issues before they result in significant financial loss or project delays.
- *Data Integration for Better Oversight:* The system integrates data from various sources, including project-specific information and publicly available data. This comprehensive approach allows for a deeper analysis of the risks associated with each project, improving the accuracy of oversight, and enabling more robust project management.
- *Resource Allocation Efficiency:* By identifying high-risk projects, Arachne helps to prioritize audit and monitoring efforts, ensuring that resources are allocated where they are most needed. This targeted approach improves the overall effectiveness of fund management and reduces wastage, thereby optimizing the use of EU structural funds.

Saler

Saler (Aa., Generalitat Valenciana, s.d.) represents The Rapid Alert System of the *General Inspectorate of Services of the Generalitat Valenciana* is in charge of spotting any anomalies in administrative files. Public procurement, grants, and fixed cash all use this system. Notices found are reported to the inspectorate so that it can conduct legal analysis and take appropriate action. Key features of Saler are:

- *Electronic Public Procurement Process:* Saler is integrated with Valencia's electronic public procurement process, which digitalizes and streamlines the acquisition and contracting activities. This integration not only increases process efficiency but also enhances the transparency and traceability of transactions.
- *Preventive Early-Warning System:* By employing a preventive approach, Saler is designed to issue early warnings for potential irregularities. This feature allows authorities to address issues before they escalate into more significant problems, thus maintaining the integrity of public services.
- *Comprehensive Data Utilization:* Saler leverages numerous data sources to monitor and analyse patterns that might indicate fraud or inefficiencies. This broad data integration helps in creating a robust framework for oversight and decision-making.
- *Risk-Based Queries:* The system employs queries that are specifically defined based on identified risks or patterns of fraud detected in the past.
- *Automated Alerts:* With automation at its core, Saler generates alerts when potential issues are detected. These automated alerts are crucial for timely intervention and ensure that the relevant authorities can act quickly to investigate and mitigate risks.

Red Flags

Red Flags (Aa., Red Flags, s.d.) is a program developed by a number of Hungarian organizations to keep an eye on public procurement in that nation and supported by the European Commission. Using predefined alerts that are rated according to their severity or likelihood of indicating actual corruption, this tool analyses procurement procedures and uses an algorithm to identify those that have the highest risk of corruption. Additionally, it makes this information accessible to both citizens and government workers. Key features of Red Flags are:

- Web-based and open source.
- Daily update of the information from Tenders Electronic Daily (TED).
- Assessment of notices to produce red flags and of complimentary information on the parties involved to produce pink flags.
- Individual set-up through the use of filters and subscriptions.
- Adaptable to other countries through the use of EU notices.

Lince

Lince (Sanz & Félix, 2019) is a tool to monitor workers on sick leave and identify potential integrated fraud, Spain's Social Security employs an artificial intelligence algorithm. With the use of this predictive tool, it assesses people’s health status and predicts the likelihood of their fitness to return to work. An algorithm is used to determine which files National Institute of Social Security (INSS) medical inspectors should review first. The system flags the file as possibly fraudulent if the reinstatement is not handled.

Rozes

Rozes (Vv., 2024), founded as a University of Padua spin-off, is an innovative artificial intelligence startup that aims to secure businesses, public administration, and entrepreneurs by offering universal, quick, and easy access to information.

Rozes has created an index using AI models to assess the degree of risk associated with a business by examining accounting irregularities brought on by fraud, money laundering, false invoicing, and fraudulent bankruptcy.

The Rozes index is an AI-based rating system Artificial Intelligence that ranks companies based on the similarity of their balance sheets to those of criminal companies.

The higher the index, the more similar the company is to ones that have accounting anomalies that could be signs of fraud, money laundering, fraudulent bankruptcy, or false invoicing. It is able to identify the company's risk and abnormal behaviour with 93 percent accuracy, well in advance of any potential negative event.

Companies are compared to similar companies and their degree of risk is appropriately positioned by relating them to their reference clusters (industry/cod. Ateco, geography, revenue size) examines the roughly 1.5 million Srls, SpAs, cooperatives, and consortia that exist in Italy; simplified Srls will soon be added to this population.

Key features of Rozes are:

- *The bankruptcy index:* The bankruptcy index, which is available to all Italian businesses that submit financial statements—including simplified or shortened versions—indicates a company's likelihood of declaring bankruptcy with a 24-month predictive capacity. It demonstrates a very high accuracy by combining mathematical analysis with advanced Artificial Intelligence methods. The index is not a "black box"; the balancing factors that went into its computation are always readily accessible, and there is a consulting service available to assess the evidence.
- *Network Analysis:* The tool makes it possible to examine and rebuild the interactive network of businesses and people in order to find instances of supply chains that pose a risk, front men, cartels, and interlocking. Potential risk chains and instances of single centres of interest in tenders and contracts can be identified thanks to the easily navigable graphical reconstruction of the network of participating companies and individuals. The algorithms of the solution analyse and categorize participating networks according to risk; the

analysis can even begin with an individual's tax code, reconstructing all of the offices and holdings of the person under analysis. With access to client or service-provided lists (PEP, crime, bad press, blacklists of various kinds, etc.), it is feasible to conduct integrated anti-money laundering and anti-terrorism checks for each of the individuals identified as being a part of the supply chain. In order to facilitate comparative analysis, all components should be aggregated into portfolios (e.g., by business area, territory, or based on any monitoring needs).

Why should LEAs use FALCON?

FALCON offers a broader and more comprehensive approach to tackling corruption compared to other tools like VigIA, ProZorro, Digiwhist, and VeriPol. While many of these existing tools focus on specific domains such as public procurement or fraud detection, FALCON is designed to manage the entire lifecycle of corruption intelligence, from data collection and analysis to informed policymaking.

It provides objective, data-driven indicators of corruption that go beyond the often-subjective measures used by other platforms. Additionally, FALCON’s versatility across multiple corruption domains, including border control and sanction evasion, makes it applicable to various corruption schemes, unlike more narrowly focused tools.

The integration of advanced AI-driven analytics and data pipelines also ensures that FALCON offers real-time insights and assessments, adhering to the principles of Trustworthy AI to ensure ethical and fair decision-making. Furthermore, its multi-stakeholder approach, engaging law enforcement, financial institutions, and policymakers, promotes collaboration and adaptability in the fight against corruption. Tailored to European challenges and compliant with emerging regulations like the AI Act, FALCON is uniquely positioned to provide LEAs with a powerful, adaptable, and forward-looking tool for combatting corruption.

2.4.3. Barriers to the adoption of AI tools to fight corruption

The implementation of new technology necessitates a thorough evaluation, considering who will be utilizing the technology, where it will be deployed, when it will be introduced, and how it will be managed. Introducing artificial intelligence tools in the fight against fraud and corruption can have several significant impacts during the change management phase:

- **Resistance to Change and Lack of Training:** Employee resistance to change is a significant barrier to the adoption of artificial intelligence in public administration. Many professionals may feel threatened by these technologies, fearing that AI will replace their jobs or disrupt their usual workflows. This resistance is often exacerbated by a lack of specialized training in using AI tools. To overcome this obstacle, it is crucial to invest in training and capacity-building initiatives that help employees understand artificial intelligence, its functions, and how it can enhance rather than replace human work. Properly managing expectations and clearly communicating the objectives and benefits of these initiatives are essential to gaining employee buy-in and facilitating a smoother transition.

- **Privacy and data protection issues:** Artificial intelligence relies heavily on the availability of large datasets to function effectively. In the public sector, handling this data raises significant privacy and personal information protection concerns. Misuse or accidental disclosure of such data can have severe consequences, particularly for whistleblowers. To ensure that AI solutions comply with current data protection regulations, it is crucial to implement robust security protocols.

- **Lack of technological infrastructure:** Artificial intelligence systems require a sophisticated and modern technological infrastructure for implementation. Therefore, it is essential to establish cooperative networks and engage national and European organizations that support the development and deployment of this technology.

- **Prejudices:** Artificial intelligence, while useful for identifying patterns and anomalies, is not infallible. Algorithms rely on data, and if this data is inaccurate or unrepresentative, the results can be unfair or incorrect. This is particularly concerning in public sectors where decisions based on faulty data can negatively impact citizens. To mitigate this risk, it is crucial to review and validate AI models critically and continuously. Additionally, protocols must be established to protect citizens' privacy and fundamental rights when using big data, allow for the swift correction of errors, and ensure that systems do not reinforce or amplify existing biases present in society or data (government automated data and systems).

- **Regulatory Compliance and the AI Act:** In addition to technical and operational barriers, legal and regulatory frameworks are becoming increasingly significant. The European Union’s AI Act seeks to regulate AI systems based on their level of risk. It introduces strict requirements for high-risk AI applications, which include many used in the public sector and in fighting corruption. These systems must meet standards for transparency, accountability, and human oversight. This regulatory environment, while essential for ensuring the ethical and safe use of AI, could present challenges for public administrations trying to integrate AI tools, as they will need to ensure compliance with these new requirements. Failure to comply could result in fines and hinder the adoption of AI-driven anti-corruption solutions. Therefore, close collaboration with legal and regulatory experts will be crucial to ensuring both ethical AI use and alignment with the upcoming legislation.

2.5. Tools and technologies to prevent and detect corruption in EU

In recent years, advancements in technology and digitization have notably enhanced the efficiency and efficacy of both public and private sectors. These developments have opened up new avenues for improving accessibility to information, fostering citizen engagement, and facilitating connectivity among people. Across the European Union, digital technologies are increasingly leveraged, serving as formidable allies in the pursuit of transparency, integrity, and accountability. Moreover, they play a pivotal role in bolstering civic participation, thereby contributing significantly to the prevention of corruption.

Today, ICT technologies should be seen as a novel component of institutional endeavours, capable of generating entirely innovative models and methodologies, rather than merely serving as tools to transition traditional analogue institutional processes and models into digital ones. While digitization encompasses a wide array of aspects, members of the NCPA have furnished examples illustrating the utilization of new digital technologies. Consequently, their contributions within the EU have led to the classification of emerging policies and practices into the following topics:

- Online training and e-learning (France).
- Digitalization of processes, data flows and management (Greece; Slovakia).
- Inter-operability of database for investigation and data gathering (Italy).
- Open data, big data, statistical analysis, and indicators (Italy).

2.5.1. Online training and e-learning

Utilizing ICT presents an initial opportunity to disseminate knowledge on anti-corruption matters. In France, combating this issue involves focusing on the "cultural aspect" of anti-corruption policies and practices, which is deemed equally significant as adhering to standards, regulations, and procedures, in line with international best practices. The primary step towards instilling this culture involves training officials directly involved in anti-corruption regulation and supervision, along with engaging all relevant stakeholders.

To address this need, the French Anti-Corruption Agency (AFA) collaborated with the National Center for Local Civil Service (CNFPT) to develop an open and extensive online course (MOOC) (Aa., MOOC, s.d.) on corruption prevention. This initiative aims to overcome the challenge of disseminating integrity standards to a widespread audience of elected officials and civil servants dispersed across the country. Upon completion of the MOOC, participants gain complimentary access to all course materials, which they can download and utilize for spreading awareness. Additionally, they can conduct in-person or virtual training sessions within their respective organizations, thereby fostering a culture of integrity and accountability at various levels.

2.5.2. Digitalization of processes, data flows and management

Over the years, there has been an increasing demand for digitization in both anti-corruption practices and institutional culture, as well as among users such as citizens, stakeholders, and activists. All aspects requiring improvement stem from the fundamental necessity of digitizing services and processes, which is essential for both present-day operations and future endeavours. However, achieving this goal necessitates not only translating analogue models into digital ones but also fostering innovation to serve as a transformative force or "game changer."

In **Greece**, to improve operational capacity and promote the organization's strategic goals, the National Transparency Authority (henceforth NTA, or Authority) has created an integrated anti-corruption strategy that makes use of technological solutions. The National Anti-Corruption Action Plan (NACAP) and the Strategic Plan of the National Transparency Authority place a high priority on advancing technological innovation and modernizing digital networks and infrastructure through the use of electronic tools and applications. Greece completely digitalized the audit function with the launch and implementation of an integrated system for the electronic document management system (hereinafter, DMS (Aa., Wikipedia.org, s.d.)). The DMS, which is connected to NTA's online complaint platform, has a unique audit workflow for the electronic management of every single case during the audit cycle, the electronic monitoring of adherence to audit recommendations, and the extraction of both quantitative and qualitative data as well as metadata, all of which are important contributions to future audit planning.

Another notable example of digitization of traditional anti-corruption tools comes from **Slovakia**. Data collection through questionnaires assessed the organization's exposure to corruption and determined which corruption risks could compromise the government office's ability to operate effectively. General corruption risk management, which includes the questionnaire, is an essential component of the anti-corruption system and an objective tool based on evidence (facts). It enables the relevant authorities to reduce the likelihood of corruption and increase the effectiveness of anti-corruption efforts. Through the use of corruption risk management, each sectoral entity determines the specific areas, positions, activities, and processes within the organization that are at risk. This is a step in the data collection process to support the creation and implementation of sensible policies aimed at reducing and eliminating these risks within the sectoral organization and its divisions. This model is based on a computer platform that offers respondents the opportunity to answer questions anonymously and quickly, simply by clicking on their chosen answer. Managers are able to promptly address corruption risks and save a significant amount of time when processing the obtained data. The program analyses the questionnaire automatically, considering the risks of corruption that are found. It might also suggest the actions to be taken based on the evaluation. The electronic application for corruption risk management, an IT tool, makes it possible to categorize and identify variables that affect the emergence and persistence of corruption risks in specific state administrative institutions.

2.5.3. Inter-operability of database for investigation and data gathering

This subject is closely tied to two core principles: standardization and simplification. Standardization focuses on integrating disparate databases to facilitate the effective use of existing knowledge, while simplification seeks to improve accessibility to this information for public administrations. The goal is to reengineer data sources and flows to streamline data collection without adding extra workloads, making the most of existing data sources to enrich institutional information.

Several EU-funded projects have been pivotal in enhancing anti-corruption efforts by tackling the challenge of database interoperability. Two prominent examples are Datacros (Aa., Transcrime, s.d.) and Kleptotrace (Aa., Kleptotrace, 2024).

Datacros is designed to detect irregularities in business ownership structures. It consolidates data from a variety of sources, including company registries, blacklists, and politically exposed persons (PEPs) records. By analyzing ownership links to high-risk jurisdictions and opaque corporate vehicles, Datacros identifies potential risks associated with money laundering, collusion, and corruption. This tool is instrumental in illuminating connections between firms, especially in high-stakes contexts like procurement processes, and in assessing complex ownership structures within multinational corporations.

In contrast, Kleptotrace focuses on combating high-level corruption and facilitating asset recovery. Kleptotrace extends the capabilities of Datacros by integrating additional asset registers (e.g., real estate, vessel ownership) and sanctions lists. This helps trace illicit proceeds and uncover sanction circumventions in transnational corruption schemes. The combination of these two tools provides law enforcement agencies and public authorities with powerful resources to tackle financial crime on a global scale.

However, one of the key challenges in integrating databases across these projects is ensuring data standardization. With data originating from different sectors and regions, it is crucial to align and harmonize diverse datasets for meaningful analysis. For example, Datacros employs over 30 risk indicators across eight dimensions, facilitating investigations into cross-border crimes by law enforcement agencies.

Both Datacros and Kleptotrace demonstrate how overcoming challenges in database interoperability—through standardized risk indicators and combined data sources—can enhance the detection and investigation of corruption and financial crime. These tools not only support police investigations and judicial authorities, particularly in transnational cases, but also aid national and local government agencies in assessing procurement collusion risks. They provide investigative journalists, non-governmental organizations, and civil society with platforms to scrutinize questionable business-political interactions.

In Italy, for instance, database interoperability is considered a key factor in measuring corruption. There is widespread belief that integrating the data held by the anticorruption authority (ANAC) with socioeconomic data managed by other public entities could create a robust knowledge base. This could be essential for assessing corruption risks and identifying early warning signs, enabling authorities to flag potential red flags before issues escalate.

These examples illustrate how the effective integration and standardization of diverse databases can strengthen anti-corruption initiatives across the EU and beyond, leveraging data-driven approaches to investigate and address financial misconduct at various levels.

2.5.4. Open data, big data, statistical analysis and indicators

The legal frameworks governing anti-corruption in various nations typically mandate the disclosure of data in an open format, emphasizing transparency of public data and information as a fundamental aspect of anti-corruption practices and policies.

To harness the potential of big data and conduct data mining analysis, statistical analysis, and the development of indicators in the anti-corruption realm, policies and practices must address the issues of open data production and release alongside the interoperability of data sources.

These principles underpin an initiative coordinated by the Anticorruption Authority (ANAC) in Italy. Known as the "Measuring the Risk of Corruption at the Territorial Level and Promoting Transparency" project, ANAC has been overseeing this initiative for an extended period.

The major goals are to support prevention and integrity, create transparency in public administration operations, and provide sufficient indicators to identify corruption at the territorial level. Three pillars support the measurement of corruption risk:

1. **Collecting** information from national databases to feed a business intelligence system that can produce dashboards with warning signs and indicators about different facets of malfeasance and corruption.
2. **Developing** a set of context and social capital indicators, as well as a set of risk indicators at the territorial level, to aid in the validation of the risk indicators, improve their interpretation, and point out potential correlations.
3. **Fortifying** an **integrity** culture, encouraging civic engagement, and disseminating information on corruption risks and the approaches used to develop and verify them.

The National Database of Public Contracts (BDNCP) (Aa., ANAC, s.d.) is the primary data source used by ANAC. A database called the BDNCP is used to gather, combine, and balance information provided by contracting authorities about public contracts. The system is compatible with other administrations' comparable systems as well as internal Authority systems. While the creation of databases and the open-access publication of information on public procurement systems are by no means new in the world, the BDNCP is notable for the breadth of its coverage as well as the volume and calibre of its data.

Broadly speaking, ANAC continues to fund the BDNCP and the public's use of its contents. It also collaborates with both institutional and non-institutional stakeholders to enhance the consistency and transparency of data releases.

The goal of "Measuring the Risk of Corruption at Territorial Level and Promoting Transparency" is to create a set of quantitative indicators that will enable integrity and transparency in public administration operations while also detecting corruption at the territorial level. Although these corruption risk indicators do not show proof of corruption, they do notify the authorities of its potential existence and enable the implementation of preventive and law enforcement measures. Beginning with the creation of indicators of corruption risks in public procurement, the project made use of the enormous amount of historical data found in the BDNCP, the ANAC National Public Procurement Database.

A few significant indicators emerged, among them:

- The number of contracts awarded based on the most economically advantageous offer.
- The proportion of negotiated procedures to open procedures.
- The value of the negotiated procedures.
- The number of contracts modified after they were awarded.
- Variations in cost and schedule at the time of execution; single bids.
- The exclusion of all bids but one.
- A high proportion of contracts for a single contractor.
- The short time between publication and the offer submission deadline.
- The duration of the tender evaluation period.

The choice to "open" the computer code used to analyse the data as well as the dataset is another creative feature. *R* is a widely used open-source software that is free of cost for statistical analysis of BDNCP and other data sources, as well as for the calculation of indicators. It covers a broad spectrum of requirements for statistical data analysis, such as the use of so-called artificial intelligence techniques and the analysis of big databases, as with the BDNCP. Moreover, "R" (Aa., R Project, s.d.) is freely available software that is open source and covered by the GPL license. Actually, in the majority of statistical research, "R" is the election software.

In summary, the global fight against corruption is evolving with the integration of advanced technologies and innovative strategies across legal frameworks and institutional practices. Initiatives such as Datacros, supported by the European Union, exemplify this trend by developing tools to detect irregularities in business ownership structures, enhancing transparency, and mitigating corruption-related risks. Moreover, the imperative for open data production and interoperability of data sources emerges as fundamental elements in facilitating data-driven analyses and strengthening anti-corruption efforts worldwide. The longstanding initiatives like the one coordinated by the Anticorruption Authority (ANAC) in Italy underscore the importance of transparent data disclosure and interoperability in effectively combating corruption. Embracing technological advancements and promoting transparency can empower institutions globally to enhance their anti-corruption capabilities, uphold integrity, and foster public trust in governance systems.

2.6. Final considerations

The FALCON project targets corruption in four key areas: bribery schemes at border crossings, penalty evasion by kleptocrats and oligarchs, fraud in government procurement, and conflicts of interest involving politically exposed persons (PEPs).

During its development and throughout its AI technology-based lifecycle, it should **follow** these **three crucial points**:

1. **The vast possibilities offered by artificial intelligence:** Artificial intelligence can identify patterns and anomalies that would be nearly invisible to the human eye or that would take a great deal of time and effort to identify because of its ability to analyse large volumes of data quickly. This not only expedites investigations and enhances the precision of identifying corrupt activities, but it can also take preventative action by detecting risk areas prior to their manifestation.

2. **Implementation must be done with caution and ethics:** Applying artificial intelligence to the fight against corruption presents several challenges. Special attention must be given to issues such as reliance on third-party vendors, algorithmic bias, and data privacy. To ensure that AI systems are fair, unbiased, and respect citizens' fundamental rights, it is essential for governments combating fraud and corruption to adopt an ethical and transparent approach. Training and raising awareness among public administration staff about these issues is crucial for the successful and ethical implementation of AI solutions.

3. **AI should be used in support of human intelligence, not in place of it:** Artificial intelligence is not a replacement for existing control and surveillance systems and mechanisms, despite its potential as an ally in the fight against corruption. Instead, it should be seen as a supplementary tool that enhances and supports administrative actions in this area. The combination of AI's analytical capabilities and human expertise offers an optimal approach for combating corruption and ensuring integrity and transparency.

3. FALCON WP5 Tools

In this section, we present the suite of tools developed under Work Package 5 (WP5) and the associated pipeline that integrates these tools to facilitate decision-making processes. WP5 focuses on Risk Assessment, Investigation, and Decision Support toolsets, harnessing advanced analytical methods and robust backend support to provide comprehensive insights and assessments. The tools developed in this Work Package are designed to leverage outputs from preceding Work Packages, particularly WP4, and integrate with a well-established knowledge base to enhance their functionality and accuracy.

3.1. Architecture of WP5

The WP5 pipeline is a structured workflow that ensures seamless interaction between various components, ultimately leading to informed and reliable decision-making. Figure shows the main components.

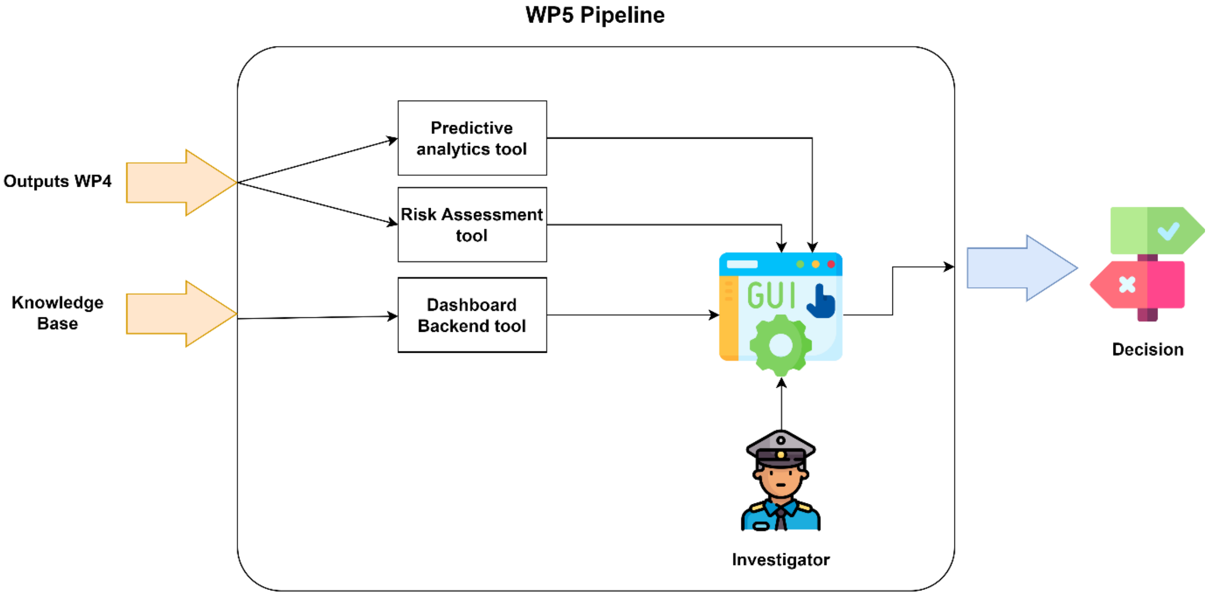

Figure 5 WP5 tools pipeline

Key elements of this pipeline include:

- **Inputs**:
  - *Outputs WP4*: These are the outputs from a previous Work Package (WP4).
  - *Knowledge Base*: This provides relevant knowledge and data to the tools in WP5.
- **Tools**:
  - *Predictive Analytics Tool*: This tool uses the inputs to perform predictive analysis.
  - *Risk Assessment Tool*: This tool assesses risks based on the inputs.

  - *Dashboard Backend Tool*: This tool supports the backend operations of the dashboard.
- **Integration and Interface**:
  - All three tools feed their outputs into a Graphical User Interface (GUI).
  - The GUI is the interface through which the investigator interacts with the tools.
- **User Interaction:**
  - *Investigator*: The investigator uses the GUI to interact with the tools, analyse the outputs, and make informed decisions.
- **Outcome**:
  - The result of this process is a decision or action to be taken by the investigator according to the information visualized.

The overall process starts with inputs from WP4 and the Knowledge Base, which are processed by the tools within WP5. The outputs from these tools are then integrated into a GUI, used by the investigator to make a final decision.

The subsequent subsections will delve deeper into each component of the WP5 pipeline, explaining their functions, interactions, and the overall impact on the decision-making process. This detailed overview aims to provide a clear understanding of how the WP5 tools are integrated and utilized, showcasing their contributions to achieving the objectives of the project.

It is important to note that this deliverable represents version 1.0 of a set of three deliverables, where each subsequent version will provide more detailed information. Additionally, certain aspects of the architecture or minor details may change in future deliverables due to the ongoing development process.

3.2. Risk assessment for corruption cases

3.2.1. Introduction to Risk Assessment in corruption cases

Corruption remains a pervasive issue that affects societies globally, leading to significant economic, social, and political consequences. The effective identification and mitigation of corruption risks are essential for maintaining transparency and integrity within organizations. A systematic risk assessment framework helps to identify weaknesses in governance structures that may be exploited by corrupt actors.

Risk assessment in corruption cases is a strategic tool used to evaluate the potential vulnerabilities in systems, processes, or organizations that may lead to corrupt practices. This assessment focuses not only on detecting existing corruption but on assessing conditions that may enable corruption in the future.

There are two principal approaches to corruption risk assessment: *qualitative* and *quantitative*.

- **Qualitative Risk Assessment**: This method involves gathering insights through stakeholder interviews, expert opinions, and perceptions of corruption risks. It is valuable for understanding institutional cultures and identifying areas with limited transparency or weak internal controls.
- **Quantitative Risk Assessment**: This approach relies on measurable data and indicators. Statistical analysis and predictive models help to identify patterns and anomalies that suggest heightened corruption risks. Quantitative tools often provide objective, data-driven insights into high-risk areas and can be more precise in detecting risks in real time.

A robust corruption risk assessment process involves several key steps:

1. **Identifying Risk Factors**: Risk factors are elements within an organization or system that may contribute to corrupt behaviour, such as the presence of discretionary power, lack of oversight, or weak enforcement of policies.
2. **Measuring Risk Exposure**: Once risks are identified, assessing the level of exposure involves analysing governance structures, compliance levels, and operational procedures to determine how vulnerable a system is to corruption.
3. **Prioritizing Risks**: Not all risks are equal. After identifying and measuring them, it's crucial to prioritize risks based on their potential impact and likelihood of occurrence. This helps allocate resources efficiently.
4. **Mitigation Strategies**: Implementing strategies to reduce or eliminate risks is the final step. Strategies may include strengthening transparency, enhancing internal controls, and using advanced tools to monitor high-risk activities.

In summary, risk assessment in corruption cases serves as a critical tool for identifying, measuring, and mitigating vulnerabilities within organizations and systems. By utilizing both qualitative and quantitative approaches, corruption risk assessment helps pinpoint areas where oversight is weak, discretionary power is high, and transparency is lacking. These assessments provide a strategic foundation for implementing targeted anti-corruption measures, ultimately enhancing accountability and fostering an environment of integrity. As corruption risks continue to evolve with advancements in technology and global interconnectedness, developing and refining robust risk assessment methodologies will remain a key component of effective anti-corruption strategies.

3.2.2. Related works

Corruption is a pervasive issue that affects economies, governments, and societies worldwide. Over the past decade, significant strides have been made in utilizing Information and Communication Technologies (ICT) to detect, prevent, and manage corruption risks. Numerous technological innovations have been introduced, combining big data analytics, machine learning, and blockchain technologies to improve the transparency and efficiency of anti-corruption efforts. These innovations have laid the foundation for modern corruption risk assessment tools like ACRA (Advanced Corruption Risk Assessment), developed under the FALCON project.

One notable development in anti-corruption technology is the application of blockchain and smart contract technologies. Benítez-Martínez et al. (Benítez-Martínez, Romero-Frías, & Hurtado-Torres, Jan 2023) introduced a governance model based on neural blockchain technology, which hosts public and private data in a decentralized manner. This system improves transparency and immutability by creating a permanent and tamper-proof record of procurement transactions, thereby reducing the potential for corrupt activities in public procurement. Similarly, Weingärtner et al. (Weingärtner, Batista, Köchli, & Voutat, 2021) employed smart contracts to automate processes in public procurement, such as bidding, supplier compliance verification, and delivery. Their work showcased how smart contracts could minimize human intervention and enhance fairness, mitigating corruption risks related to hidden agreements or biased bidding.

In addition to these contributions, Mazza et al. (Mazza, Cappellari, & Renzi, 2022) explored a blockchain-based approach to enhancing transparency in public procurement processes. Their study demonstrated that blockchain could significantly improve the audibility and integrity of procurement transactions, helping to eliminate opportunities for manipulation or fraud. This further strengthens the argument for blockchain as a vital tool in reducing corruption risks in governmental procurement systems.

The application of machine learning (ML) to detect corruption is another significant advancement in the field**.** Caruso et al. (Caruso, Bruccoleri, Pietrosi, & Scaccianoce, May 2023) developed a system that leverages supervised neural networks to monitor public procurement processes, identifying suspicious transactions and irregular patterns. This approach addresses the issue of KPI overload by reducing the number of key performance indicators (KPIs) needed to monitor large datasets, while still ensuring a low probability of missing important warnings. Machine learning’s ability to process vast amounts of data and detect hidden patterns has made it a critical component in modern anti-corruption efforts.

Similarly, in (Dawson & Rahman, 2022) , the authors examined the role of artificial intelligence (AI) and predictive analytics in enhancing anti-corruption efforts. Their research emphasizes how law enforcement can leverage AI to predict corrupt activities before they occur, allowing authorities to focus on prevention rather than mere detection. By analysing historical data, AI systems can detect early warning signs of corruption, enabling proactive interventions.

In (Mungiu-Pippidi, 2021), the importance of transparency and data-driven approaches in curbing corruption across the European Union is highlighted. The study delves into corruption control policies and how the EU has implemented digital transparency tools to foster greater accountability in government operations. The work illustrates how transparency initiatives, coupled with digital tools, reduce opportunities for corruption by enabling public oversight and real-time monitoring of governmental transactions.

In the same vein, Wacker et al. (Wacker, Ferreira, & Ladeira, 2018) introduced a system that utilizes convolutional neural networks (CNNs) to analyse images of supplier locations from Google Street View, distinguishing between legitimate businesses and fake suppliers. This innovative approach helps authorities detect fraudulent suppliers, a common issue in procurement corruption.

Fuzzy logic has proven to be an effective tool for dealing with the uncertainty and qualitative aspects of corruption risk. By simulating human reasoning, fuzzy logic systems can analyse imprecise data and provide meaningful assessments. Gallab et al. (Gallab, Bouloiz, Alaoui, & Tkiouat, 2019) demonstrated the effectiveness of fuzzy logic in risk assessment by calculating Risk Priority Numbers (RPNs), allowing for the evaluation of risk based on severity, frequency, and detectability. These systems provide the flexibility needed to assess risks in environments where data may be incomplete or inconsistent, making them particularly useful in corruption risk assessment.

Building on these technological advancements, the ACRA tool aims to provide a comprehensive platform for real-time corruption risk assessment and investigation prioritization. FALCON is currently being developed in the context of the FALCON project, aiming to integrate advanced analytics to process and analyse diverse data sources, including financial transactions, procurement records, and cryptocurrency data. By using a combination of quantitative and qualitative methodologies, ACRA shall be able to identify potential corruption risks by analysing historical trends and real-time data streams.

ACRA's architecture is designed to facilitate anomaly detection and risk scoring by leveraging a range of predefined corruption risk indicators, such as discretionary authority, transparency, and conflict of interest. ACRA aims to empower Law

Enforcement Agencies (LEAs) and Anti-Corruption Authorities to make informed decisions based on the likelihood and impact of corruption risks.

3.2.3. Introduction to the Advanced Corruption Risk Assessment (ACRA) Tool

In the ongoing battle against corruption, technological advancements have provided new avenues for law enforcement and anti-corruption bodies to detect, manage, and prevent corrupt practices. ACRA is designed to offer comprehensive, data-driven risk assessments in real time, enabling Law Enforcement Agencies (LEAs) and Anti-Corruption Authorities to identify high-risk corruption cases with enhanced accuracy and efficiency.

ACRA will be able to detect patterns and anomalies that indicate potential corruption risks, such as irregularities in procurement, company data or financial transactions. It will offer several advanced features including:

1. *A Real-Time Monitoring and Alert System:* One of ACRA's most powerful features will be its real-time monitoring capability. ACRA will continuously scan incoming data streams for signs of corruption, flagging suspicious activities or anomalies as they occur. The automated alert system will ensure that anti-corruption bodies are notified when high-risk events arise. Moreover, ACRA's real-time capabilities will be enabled by its ability to process large volumes of data in real time, flagging irregularities as they occur.
2. *Fuzzy Logic for Qualitative Assessments:* In addition to its quantitative capabilities, ACRA will integrate fuzzy logic to account for subjective assessments of corruption risks, such as the discretionary authority of officials or the level of transparency in specific processes. This will allow the system to evaluate qualitative factors that are often critical in understanding corruption risks but difficult to measure precisely.

The architecture of the ACRA tool is both robust and flexible, allowing for the integration of multiple data streams and supporting real-time data processing. The system's architecture is divided into three core components:

1. **Backend Service Layer**: This layer is responsible for processing and analysing the diverse data streams fed into the system. ACRA is built on a Flask framework, which allows for the efficient handling of large-scale data, including structured datasets like financial records and unstructured data such as free-text reports. The backend ensures that all incoming data is analysed and cross-referenced to produce meaningful insights into potential corruption risks.
2. **RESTful Service Layer**: The REST API will act as the intermediary layer, managing communication between the backend services and the user-facing

interface. This service is responsible for ensuring that data flows seamlessly between the processing components and the end users, who interact with ACRA through the graphical interface. By leveraging RESTful services, ACRA will be able to provide real-time responses and dynamic data updates, which are crucial for monitoring ongoing corruption risks.

3. **Graphical User Interface (GUI)**: The frontend of ACRA will be designed with user experience in mind. It will provide a visual dashboard where users can interact with risk indicators, analyse trends, and prioritize cases based on their potential for corruption. The GUI will allow users to view historical data, monitor real-time events, and receive alerts for emerging risks. It will also support the customization of risk parameters, enabling users to tailor the tool to specific sectors or types of corruption.

The **Advanced Corruption Risk Assessment (ACRA)** tool will be built upon a robust and modular architecture designed to facilitate real-time corruption risk assessment and analysis. Its architecture will integrate multiple components that work together seamlessly to process large datasets, analyse corruption indicators, and deliver actionable insights through an intuitive user interface. The core architecture of ACRA is composed of three primary elements: **Data Input & Corruption Indicators**, the **ACRA Core**, and the **ACRA User Interface**.

A. Data Input & Corruption Indicators

One of the core parts of the ACRA tool will be its data input system, which will gather information from a wide array of data sources. These data sources will include public procurement records, financial transactions, textual reports, cross-border financial flows, and cryptocurrency data, etc. By aggregating and standardizing data from these diverse inputs, ACRA will ensure that it has comprehensive coverage of the key sectors where corruption is most likely to occur.

In addition to raw data, the ACRA tool will incorporate corruption indicators that are predefined based on research and expert input. These indicators are vital in guiding the system's analysis, providing a framework for detecting suspicious patterns or anomalies that may be indicative of corrupt behaviour.

B. ACRA Core Mechanisms

The heart of the ACRA architecture is the **ACRA Core**, which will be responsible for processing and analysing the data received from various sources. This core will provide two main services as follows:

1. *Flask-based Backend Services:* The backend of ACRA will be built on the Flask framework, which allows for efficient and scalable data analysis. It will process large volumes of both structured and unstructured data, including financial records, procurement logs, and other transactional data. The backend services will detect anomalies, predict corruption risks, and provide actionable insights based on historical and real-time data analysis, etc.
2. *Intermediate RESTful Services:* The RESTful API will serve as the intermediary between the backend analysis services and the user interface. This service will ensure that data flows smoothly between the system's internal components and is made available in real time to end-users. It will also facilitate expert feedback, enabling domain experts to interact with the system and provide input that can refine the accuracy of the corruption risk assessments.

C. ACRA User Interface

The **ACRA User Interface (UI)** will be designed with the user experience in mind, providing an intuitive and interactive platform for law enforcement agencies and anti-corruption authorities to assess and prioritize corruption risks. The user interface will be powered by the REACT Framework**,** which ensures smooth interaction and visualization of complex data sets. Different levels of access will be provided depending on the end-user (e.g., LEAs, scientific stakeholders, technical developers, SSH experts)

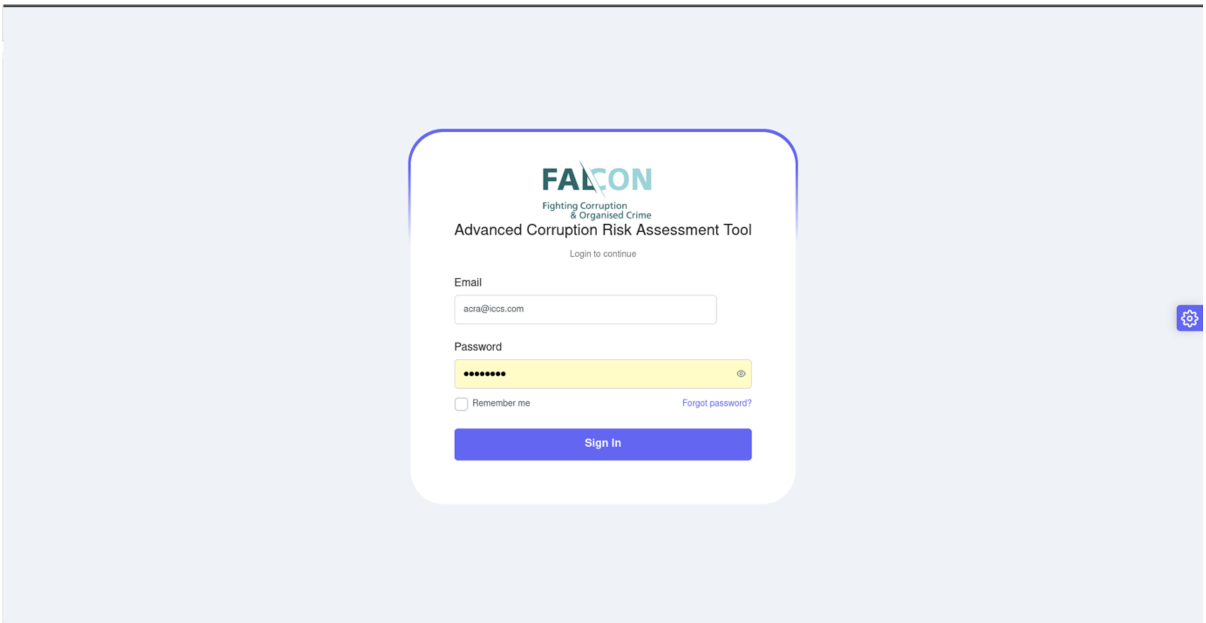

Figure 6 The ACRA tool's login page

Key features of the ACRA User Interface will include:

- **An Interactive Risk Analysis Table**: This table will allow users to explore the data collected and analysed by the system, with real-time updates on corruption risks in various sectors. This component will support pseudonymization mechanisms for the corresponding identifier IDs.

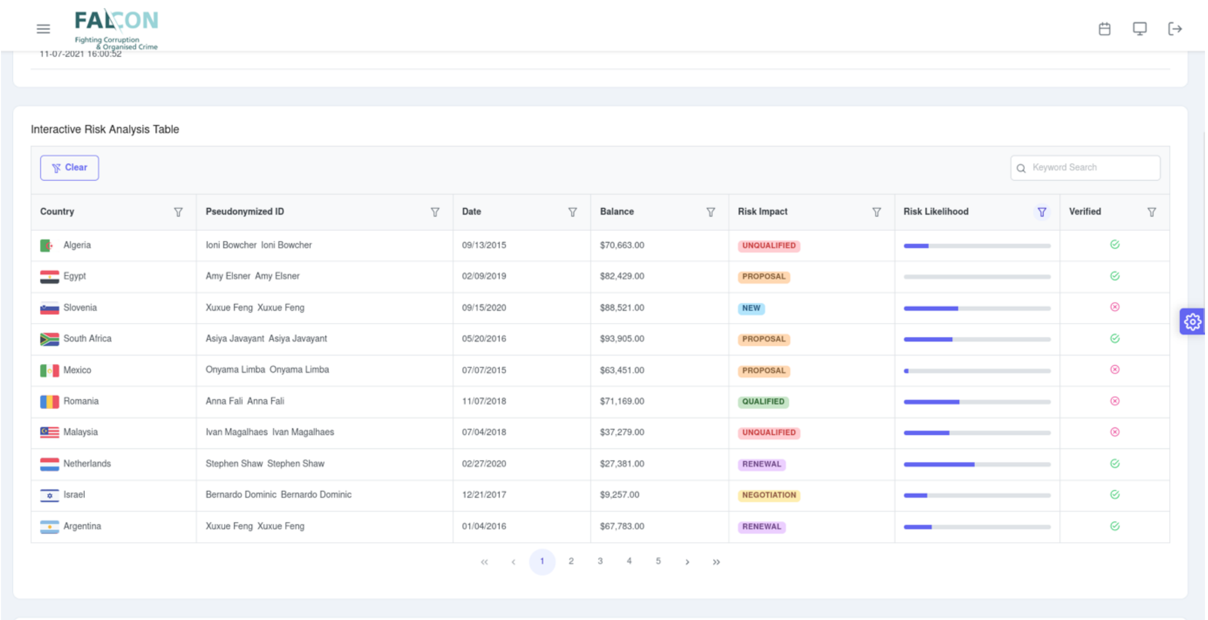

Figure 7 An Interactive Risk Analysis Table generated by ACRA

- **A Risk Matrix**: A graphical representation of the relative severity and likelihood of different corruption risks. This matrix will help users to prioritize high-risk areas and focus investigative resources where they are most needed.

Figure 8 A risk matrix generated by ACRA

- **Interactive Maps**: ACRA's map functionality will generate a geographic view of corruption risks, once ingested with the appropriate inputs. This will allow users to visualize corruption hotspots and analyse patterns based on geographic data. This feature will be particularly useful for tracking cross-border corruption and illicit financial flows.

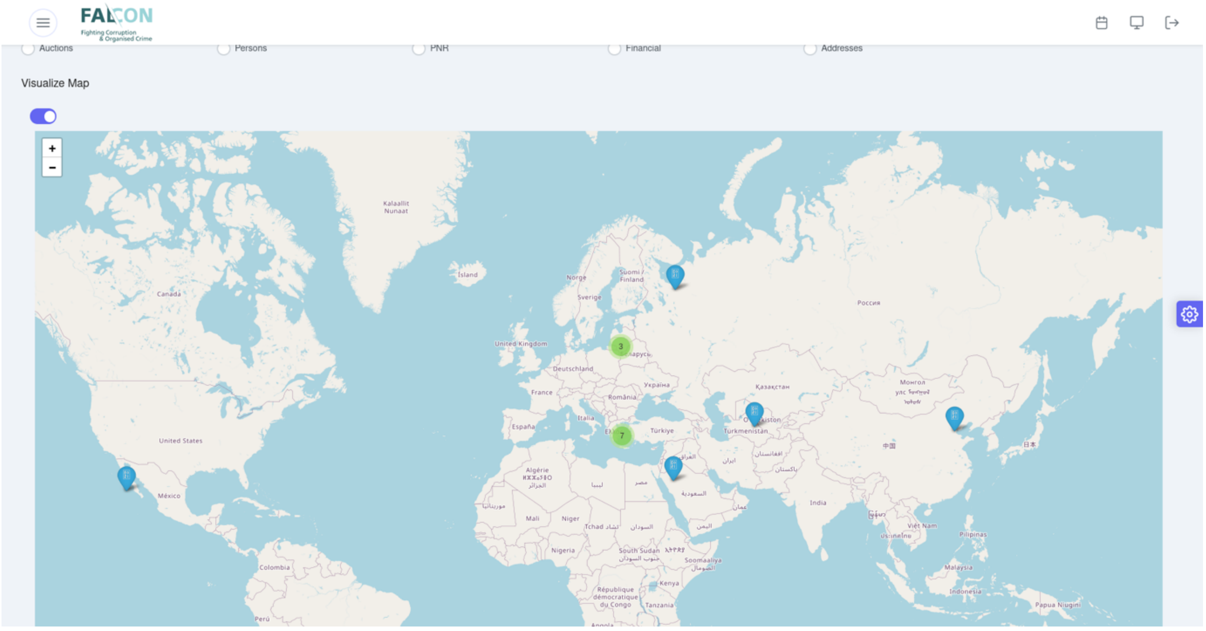

Figure 9 Interactive maps generated by ACRA

Together, these components aim to render ACRA a powerful tool for detecting and mitigating corruption risks. By leveraging advanced data analytics, machine learning, and intuitive user interfaces, the ACRA tool empowers authorities to take a proactive approach to corruption detection, ensuring that high-risk activities are flagged in real-time and investigated efficiently.

This comprehensive description details the conceptual architecture of the ACRA tool as depicted in Figure 10.

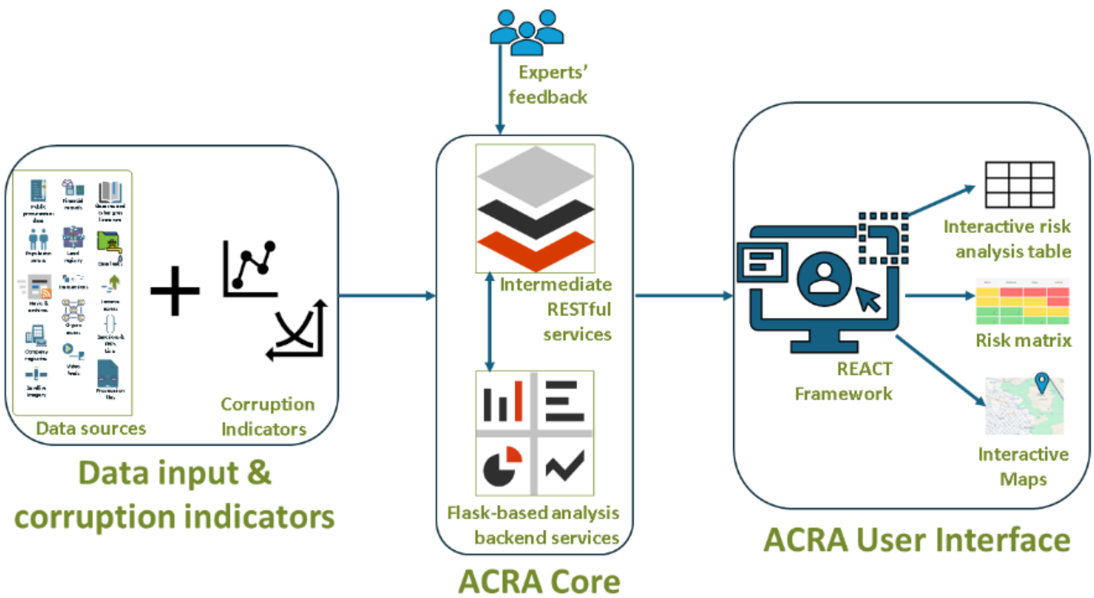

Figure 10 Advanced Corruption Risk Assessment's tool conceptual architecture

When ACRA detects suspicious activities or anomalies, the tool will generate automated alerts. These alerts will provide authorities with the opportunity to investigate potential risks before they escalate into full-scale corruption. For example, if ACRA identifies a high-value contract awarded to a company with known ties to a government official, the system will flag this event for review by anti-corruption bodies.

Real-time monitoring can be valuable in cross-border investigations, where detecting suspicious financial flows is critical to uncovering corruption. ACRA will be able to integrate data from cryptocurrency transactions and bank transfers, offering authorities a broader view of how corruption operates on a global scale.

ACRA will leverage a range of predefined risk indicators to assess and rank corruption risks across different sectors. These indicators will be based on established risk factors that have been proven to correlate with corruption. ACRA's risk indicators will allow for the quantification of corruption risks, enabling authorities to prioritize cases based on severity and likelihood.

Some of the indicators monitored by ACRA will be related to:

- **Discretionary Authority**: The degree of power an official or organization holds in making decisions without oversight is a major risk factor. ACRA will track instances where officials have disproportionate control over financial or administrative decisions, flagging these as high-risk areas.
- **Transparency and Accountability**: ACRA will monitor public procurement processes, financial reporting practices, and legislative decisions to assess the level of transparency. A lack of transparency is often a precursor to corrupt activities, as it enables decision-makers to operate without scrutiny.
- **Regulatory Complexity**: Corruption thrives in complex regulatory environments where rules are difficult to enforce. ACRA's analysis of regulatory structures will help in identifying areas where legal loopholes may be exploited by corrupt actors.
- **Conflict of Interest**: One of the clearest indicators of corruption is a conflict of interest, where personal relationships or financial interests influence decision-making. ACRA will evaluate potential conflicts of interest by cross-referencing personal and financial data with organizational roles.

ACRA will be able to leverage big data analytics, real-time data processing, and predictive algorithms to offer actionable insights that can be used to prevent and combat corruption. By integrating multiple data sources (e.g., financial records, cryptocurrency transactions, and procurement logs), ACRA will provide a holistic view of corruption risks. This multi-dimensional analysis is essential for addressing corruption's inherently complex nature.

ACRA's approach, which will be capable of combining quantitative analysis and predictive models with qualitative assessments, will ensure that all aspects of corruption risks are thoroughly evaluated. The tool is also being designed with ethics and explainability in mind. This will ensure that the decision-making process of the tool is transparent and accountable, respecting the human rights and privacy of individuals involved in anti-corruption investigations.

By automating the risk assessment process, ACRA will be capable of reducing the burden on anti-corruption bodies, enabling them to process larger volumes of data in shorter timeframes. Through its comprehensive, data-driven approach, ACRA aims to act as a valuable asset in the fight against corruption, particularly in sectors where traditional monitoring methods have struggled to keep pace with evolving risks. As corruption schemes become increasingly sophisticated, particularly in cross-border operations and financial transactions, ACRA's advanced analytics may play a crucial in detecting hidden patterns, identifying illicit financial flows, and exposing money laundering schemes linked to corruption. Moreover, as more data becomes available and ACRA continues to refine its mechanisms, its ability to predict and prevent corruption will improve.

3.2.4. ACRA Pipeline

The ACRA tool will operate through a streamlined pipeline designed to process vast amounts of data efficiently, will analyse potential corruption risks, and deliver actionable insights to anti-corruption authorities. The pipeline will consist of several stages, each contributing to the accurate assessment of corruption risks in real-time.

A. Data Collection and Integration

The pipeline will begin with data collection, where various data sources will be aggregated and integrated into the system. These data sources include:

- **Public procurement records** (e.g., contracts, bids, awards)
- **Financial transactions** (e.g., bank transfers, cryptocurrency flows)
- **Regulatory and legal documents** (e.g., laws, regulations, compliance data)
- **Cross-border financial activities** (e.g., international money flows, offshore accounts)

Each data source will provide crucial information needed to assess corruption risks. ACRA will collect data from both *structured sources* (such as databases, transaction logs) and *unstructured sources* (such as textual reports, media articles). The integration layer will ensure that data is standardized and cleaned for further processing, addressing any inconsistencies or missing data points before moving forward in the pipeline.

B. Data Preprocessing and Feature Extraction

Once the data will be collected, it will undergo preprocessing. This stage involves cleaning, normalizing, and transforming raw data into a suitable format for analysis. Feature extraction will then be applied, identifying key attributes or patterns within the data that are relevant to corruption risk assessment.

For example, in procurement data, features such as contract amounts, bidder history, and award frequency will be extracted. In financial data, features like suspicious transaction amounts, unusual account activity, and cross-border fund transfers will be isolated for further analysis. Text mining techniques may also be applied to unstructured text to extract relevant keywords or sentiments linked to corruption.

C. Risk Scoring and Prioritization

Once the data is analysed, ACRA will calculate risk scores for various entities, such as procurement contracts, companies, or government officials. The risk score will be based on multiple factors, including discretionary authority, lack of transparency, conflicts of interest, and financial irregularities. These factors will be quantified into a comprehensive risk profile for each entity or transaction.

ACRA will then prioritize cases based on the severity and likelihood of corruption. High-risk cases will be flagged for immediate attention, allowing authorities to focus their investigative resources efficiently. The tool will also provide a risk matrix to visually represent the risk scores and their corresponding impact, making it easier for users to understand and act on the insights generated by the system.

D. Visualization and Reporting

The final stage of the pipeline will be the visualization and reporting. ACRA presents the analysed data and risk scores through an intuitive dashboard, offering various visualization options such as:

- **Interactive risk analysis tables** for detailed data inspection
- **Risk matrices** to display the relationship between likelihood and impact of corruption risks
- **Geographic maps** for tracking corruption hotspots or cross-border financial activities

Users can generate customized reports based on specific risk scenarios or datasets, providing comprehensive insights for decision-making. These reports can be shared across departments or with external stakeholders to drive coordinated anti-corruption efforts.

E. Feedback Loop and Continuous Improvement

A crucial aspect of ACRA's pipeline will be the incorporation of a feedback loop. Experts and investigators will continuously provide feedback on the system's assessments and outcomes, which will be used to refine the machine learning models and improve the accuracy of the tool over time. This will ensure that ACRA remains adaptable to evolving corruption schemes and continues to provide relevant and actionable insights.

Figure 11 depicts the ACRA's pipeline schema.

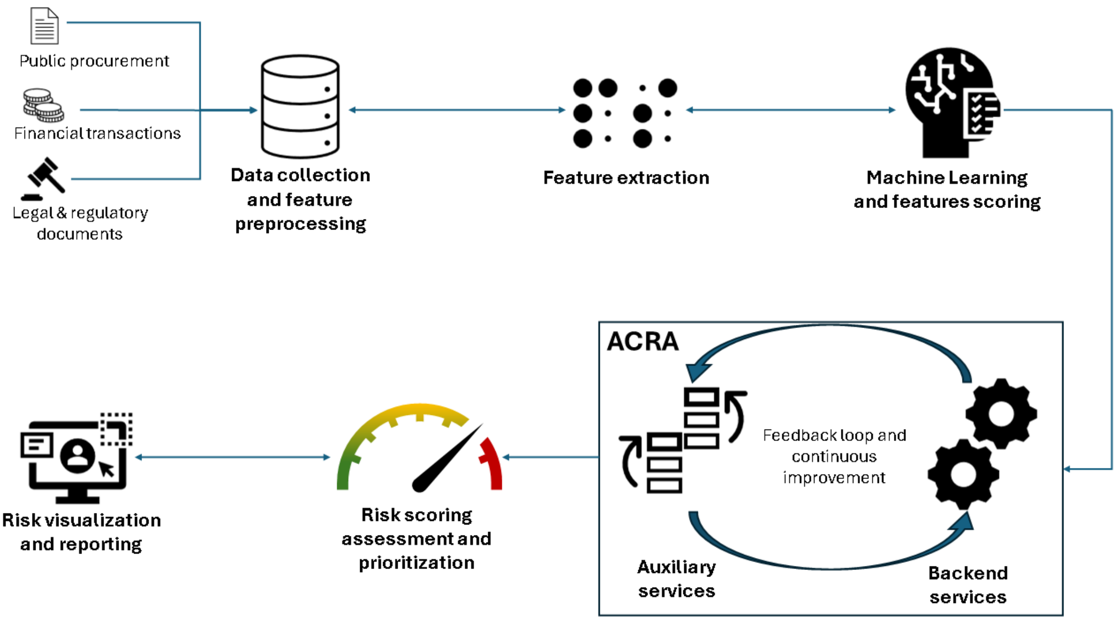

Figure 11 Advanced Corruption Risk Assessment's tool pipeline

3.2.5. Limitations and Challenges

It is important to acknowledge the potential challenges and limitations of ACRA's implementation. These considerations are crucial for ensuring that ACRA can deliver on its promise of enhancing transparency, efficiency, and effectiveness in detecting corruption.

A. Data Availability and Quality

The success of the tool is inherently dependent on the availability and quality of data. In many regions, particularly those with weaker governance structures, the collection and accessibility of reliable data can be problematic. The following issues can arise:

- **Incomplete Data**: The lack of comprehensive data sources can limit the tool's ability to assess risks accurately. Missing or inconsistent data, particularly in procurement processes or financial records, can lead to erroneous or incomplete risk assessments.

- **Data Fragmentation**: Information related to corruption is often spread across multiple agencies and platforms, which may not communicate effectively. This fragmentation can reduce the effectiveness of ACRA's real-time monitoring capabilities, as critical data sources may not be fully integrated.
- **Data Quality**: The quality of the data also varies significantly between countries or even departments. Poor data entry practices, outdated systems, and lack of proper auditing can result in inaccuracies, which in turn affect the reliability of the ACRA tool's predictions and alerts.

To mitigate these challenges, investments in data collection infrastructure, standardization of data formats, and regular audits of data quality are necessary. Additionally, collaborative efforts between government bodies and international organizations to enhance data sharing can significantly improve ACRA's functionality.

B. Technical Integration and Interoperability

ACRA's integration into existing systems may present a technical challenge, particularly for organizations that rely on legacy IT systems or have fragmented data environments. Some of the most prominent challenges include:

- **Compatibility with Legacy Systems**: Many law enforcement agencies (LEAs) and/or anti-corruption authorities still operate on older IT infrastructures, which may not be easily compatible with ACRA's real-time data processing capabilities. Upgrading or replacing such systems could require substantial investment and time.
- **Interoperability**: ACRA is being designed to aggregate data from diverse sources, including structured and unstructured datasets such as financial records, procurement logs, and cross-border transactions. Ensuring that ACRA can seamlessly integrate with these varied systems—each with its own formats, protocols, and security requirements—can be a complex task. Poor interoperability could lead to delays in data processing, reducing the effectiveness of real-time monitoring.

Addressing these issues will require tailored solutions, including developing APIs (Application Programming Interfaces) and middleware to bridge gaps between old and new systems. This process may also necessitate close collaboration with IT departments to ensure smooth integration and uninterrupted functionality.

C. Scalability and Performance

Maintaining scalability without compromising performance presents a considerable challenge. The tool aims at handling increasing volumes of data while still providing timely and accurate risk assessments.

- **Computational Demands**: Real-time data analysis, particularly across large datasets, can place a significant load on processing systems. Ensuring that ACRA remains responsive, as the volume and complexity of data grow, requires powerful computational infrastructure, which may not be available to all institutions, especially those in developing regions.
- **Resource Allocation**: The allocation of computational resources will also need to scale in proportion to the demand. There may be a need for cloud-based solutions, or distributed computing, to ensure that ACRA continues to perform efficiently as it grows. However, transitioning to cloud solutions may introduce other challenges, including concerns about data security and jurisdictional control over sensitive data.

D. User Training and Adoption

The successful implementation of ACRA depends not only on its technical capabilities but also on its adoption by end-users. This includes law enforcement agencies (LEAs), anti-corruption parts, and/or other stakeholders who may interact with the tool in a frequent basis.

- **Training Needs**: ACRA's advanced features, such as the real-time monitoring, require users to have a certain level of technical proficiency. Proper training programs will need to be developed to ensure that users can fully leverage ACRA's capabilities and interpret its results accurately.
- **User Resistance**: Like with any new technology, there may be resistance to adopting ACRA, especially if it is perceived as overly complex or if users fear it could disrupt existing workflows. Efforts to address these concerns through clear communication of the tool's benefits, as well as providing ongoing support and user feedback mechanisms, will be critical to ensuring successful adoption.

In conclusion, the **ACRA** tool's combination of **real-time data analysis, predictive analytics capabilities**, and **interactive visualization** will enable LEAs and Anti-Corruption Authorities to better understand, anticipate, and respond to corruption risks. ACRA aims to not only enhance the efficiency of anti-corruption efforts but also shift the focus from reactive investigations to proactive prevention. As the global fight against corruption intensifies, tools like ACRA will play an increasingly vital role in safeguarding transparency, integrity, and accountability across public and private sectors.

3.3. Predictive analytics for corruption cases

This task focuses on the implementation of artificial intelligence techniques to support predictive analytics, targeting mainly fraud prediction in procurement processes, as well as cryptocurrency transactions. By harnessing historical data related to specific areas of interest (such as procurement databases, cryptocurrency exchange records, etc.), and incorporating information from heterogenous sources, such as World Bank contracts, the aim is to predict fraudulent activities with a high degree of precision in near-real time, as well as to provide accurate future forecasts. Additionally, fraud prediction tools developed using supervised machine learning and deep learning technologies by FALCON partners yielding state-of-the-art outcomes. These advancements are pivotal in pre-empting procurement fraud and identifying anomalous patterns, thus ensuring the security and trustworthiness of financial engagements.

3.3.1. Introduction to Procurement Fraud Analysis

The early prediction of fraudulent activities in procurement procedures is a critical aspect of modern governance and financial management. Proactive detection (i.e., prediction) of fraud can have a profound impact on the efficiency and integrity of procurement processes, which are often vulnerable to various forms of corruption and malfeasance. Next, an introduction to its importance is outlined:

1. **Cost Savings and Resource Allocation**: Early fraud detection systems can save significant amounts of public funds by preventing the misallocation of resources. By identifying potential fraud before contracts are awarded, organizations can avoid overpayments, and the costs associated with legal proceedings and contract cancellations.
2. **Enhancing Trust and Transparency**: Proactive fraud detection enhances the transparency of the procurement process, thereby increasing trust among stakeholders. This is particularly important for public institutions, where trust is a cornerstone of legitimacy and effective governance.
3. **Deterrence of Future Fraud**: The presence of an effective early warning system can act as a deterrent to potential fraudsters. Knowing that fraudulent activities are likely to be detected early makes the risk of engaging in such activities higher, potentially reducing the overall incidence of fraud.
4. **Legal and Regulatory Compliance**: With increasing regulatory scrutiny, organizations are under pressure to demonstrate due diligence in preventing fraud. Early detection systems help in complying with these regulations and can protect organizations from penalties and reputational damage.

In conclusion, the early prediction of fraud in procurement is not just about preventing financial loss; it's about maintaining the integrity of the procurement process, ensuring fair competition, and upholding the principles of good governance. It's a proactive measure that supports the strategic objectives of organizations and contributes to the overall health of the economic system. Therefore, early prediction of fraud in procurement data is essential for mitigating risks and preventing financial losses.

Here are some approaches to predict early fraud effectively:

1. **Data Analytics and Anomaly Detection**: By analysing procurement data, anomalies that deviate from normal patterns can be identified. This includes unusual transactions, pricing discrepancies, and irregular bidding behaviour.
2. **Machine Learning Predictive Models**: These models use historical data to predict potential instances of procurement fraud. They can be trained to recognize patterns associated with fraudulent activities.
3. **Network Analysis**: Examining the relationships between entities involved in procurement can reveal hidden patterns and connections that suggest collusion or other fraudulent behaviours.

3.3.2. Related Works

In (Carneiro, Veloso, Ventura, Palumbo, & Costa, 2020), the authors present an innovative system designed to enhance fraud early detection in public procurement processes in Portugal. The system described in the paper consists of four main components:

1. **Data Acquisition**: Public procurement data and information about involved entities are collected from public sources.
2. **Graph-Oriented Database**: The collected data is integrated into a graph-oriented database, which facilitates complex network analysis.
3. **Rules-Engine**: This component enriches the data with additional information using legal rules or custom rules defined by users.
4. **Graph-Oriented User Interface**: It supports decision-making by allowing users to explore and filter information efficiently in a natural and geo-referenced manner.

The primary goal of this system is to increase transparency and contribute to the fairness of the public procurement procedure by making relevant information more accessible. The paper emphasizes the need for automated processes in fraud detection, leveraging statistical, computational methods, and ML to reduce manual effort and focus investigative attention on relevant cases.

In (Dhurandhar, Graves, Ravi, Maniachari, & Ettl, 2015), a robust tool designed to identify fraud and risk in procurement processes is proposed. The tool described in the paper analyses both standard transactional data and multiple public and private data sources, providing a comprehensive coverage of fraud types. The analytical components of the tool are adaptable and could be used for detecting fraud in domains beyond procurement. One of the key features of the tool is its learning component, which is based on feedback from investigations and has formal guarantees. The deployment of this tool over 12 months has led to the discovery of many cases of compliance risk and fraud across more than 150 countries and 65,000+ vendors. It has increased the number of true positives found by over 80% compared to other state-of-the-art tools that domain experts were previously using. This indicates a significant improvement in the ability to detect risky procurement entities, which is crucial for preventing fraud and ensuring the integrity of procurement processes.

The authors of (Bai & Qiu, 2023) explores the application of machine learning (ML) techniques to predict/detect procurement fraud, a significant issue in free markets. The authors collaborated with SF Express, which provided access to procurement data from 2015 to 2017. They developed neural network (NN) models that represent each procurement event with nine specific features to identify and classify types of procurement fraud. The models were tested on over 50,000 samples, demonstrating their effectiveness in predicting fraudulent activities. Finally, this study highlights the potential of ML in enhancing audit processes and combating one of the most profound crimes in the Chinese market.

The authors in (Aran, et al., 2020) present a significant study on the use of ML to combat corruption in public procurement processes. In particular, the paper details a partnership with the National Agency for Public Procurement (DNCP) in Paraguay. The researchers developed a system that analyses procurement data at the time of creation and assigns a risk score indicating the likelihood of future complaints. This score predicts whether a procurement is likely to receive complaints in the future, which are used as a proxy for corruption. Their model showed a substantial improvement in the complaints capture rate, increasing it from 30% to 78%, thus demonstrating a more efficient and less biased review process towards high-value procurements.

(Gallego, Rivero, & Martínez, 2018) delves into the potential of ML to forecast corruption and inefficiency in public procurement. The paper utilizes a dataset of over 2 million public contracts in Colombia to train ML models that predict which contracts may face corruption investigations or implementation inefficiencies. It addresses the challenges of class imbalance in data and reports high accuracy in the models' predictions. The authors also discuss the trade-offs between precision and recall and identify key features that contribute to predicting malfeasance within contracts.

In (Lima, et al., 2020), the authors utilized various ML and Deep Learning (DL) techniques to predict/detect fraud in public procurement. The study introduces a new dataset from the Brazilian official journal, consisting of over 15 million textual entries, including 1,907 annotated as risky for collusion and fraud. The authors employ a Bidirectional Long Short-Term Memory (Bi-LSTM) and a bottleneck DL network (consist of Dense layers with ReLU activation function), which outperform classical classifiers with a precision of 92.4% and 93.0%, respectively. This research is significant as it not only provides a novel dataset for public procurement analysis but also demonstrates the effectiveness of DL models in identifying potentially fraudulent activities in a large-scale, real-world context.

In (García Rodríguez, Rodríguez-Montequín, Ballesteros-Pérez, Love, & Signor, 2022), the authors evaluate the effectiveness of eleven ML algorithms in identifying collusion within public procurement auctions. To that end, datasets obtained from Brazil, Italy, Japan, Switzerland, and the United States were considered. These datasets contain historical auction data, where the scope of ML algorithm is to identify non-competitive bids that are indicative of collusion. Moreover, this study found that models like Random Forest and Gradient Boosting demonstrated superior performance, particularly showcasing lower false positive. This is a crucial aspect when identifying collusion in public procurement, as it minimizes the risk of incorrectly flagging innocent parties.

(Decarolis & Giorgiantonio, 2022) provides a detailed analysis of quantitative indicators, known as red flags, to predict corruption in public procurement. The study presents three main contributions:

2. **Expansion of Red Flags**: It expands the set of commonly discussed indicators in the literature to new ones derived from the operative practices of police forces and the judiciary.
3. **Validation of Indicators**: Using novel and unique data on firm-level corruption risk, the study validates the effectiveness of these indicators.
4. **Prediction Ability**: It quantifies the increased corruption-prediction ability when indicators that are unavailable to the corruption-monitoring authority are included in the prediction exercise.

The paper highlights that the most effective red flags in early detecting corruption risks are those related to discretionary mechanisms for selecting private contractors, compliance with the minimum time limit for the submission of tenders, and subcontracting.

Finally, in many papers (Ferwerda, Deleanu, & Unger, 2017), (Goryunova, Baklanov, & Ianovski, 2021), (Lyra, Damásio, Pinheiro, & Bacao, 2022) dealing with corruption within procurement public datasets, mention the problem of missing labelled data. Therefore, they consider various indicator as proxy of corruption, such as the single bidding. To that end, within FALCON project, we assume single bidding as an indicator of corruption in public procurement activities.

3.3.3. Predictive Analytics Architecture

In the proposed predictive tool, the main goal is to provide prediction of possible single bidding at the beginning of a tender process. Therefore, insights that are provided in the beginning of the tender process will be utilized to predict the whether the tender under consideration will receive a single bidding or not. Furthermore, different supervised DL models per country will be trained to provide predictions, as the tender process is different in country-level. More details on the DLs and MLs to be used and their performance will be presented in 2nd round of the deliverable: Risk Assessment, Research and Decision Support Toolkit deliverable (D5.2).

The procurement datasets data that will be considered are provided by the **Opentender.eu** dataset (Opentender., n.d.), which includes detailed information on public procurement across 35 jurisdictions, including the 28 EU member states, Norway, the EU Institutions, Iceland, Switzerland, Georgia, Serbia, and North Macedonia. It allows users to search and analyze tender data, offering insights into various sectors, regions, suppliers, buyers, and tenders.

Next, more features-variables will be considered during the deployment of the tool, such as online news for the country of interest, hence, new performances analysis will be done in D5.2. To that end the **GDELT (Global Database of Events, Language, and Tone)** database (Project, n.d.) will be considered. GDELT is a comprehensive and continually updated open-access repository of global news-media data (translated into English). It captures and analyzes news articles, broadcasts, and online sources from around the world to provide valuable insights into a wide range of events, emotions, and trends.

Next, the workflow of the proposed approach is outlined (see Figure 12):

1. Historical procurement and GDELT data for the country of interest are selected. Data is retrieved from the FALCON platform (available in KB or other sources).
2. The data are filtered, using only the features-variables related to the single bidding contracts. A ML model (e.g., Random Forest) is pretrained (per country) to define the most relevant features to the single bidding contracts.
3. The selected features are used as input to the pretrained DL-based classifier.

4. The output stands for the likelihood of the contract to receive a single bidding.

The results will be presented in a JSON format on the service and user interface tool. The required input parameters (e.g., country, contract id for consideration etc.) will be passed to the service via a POST request. Finally, it is worth mentioning that the input data used in the DL-based classifier is known in the beginning of the tender procedure. In addition, only previous tender activities (historical data) and their outcomes (single bidding or not) are utilized.

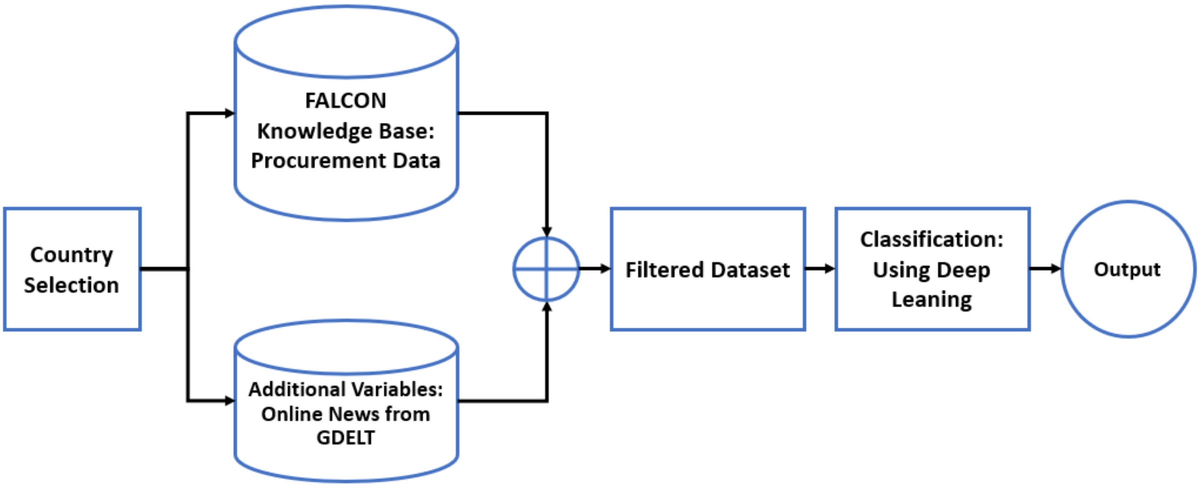

Figure 12 Basic workflow of predictive analytics tool.

Finally, the predictive analytics tool will leverage publicly available datasets (e.g., from OpenTender (Aa., Opentender, 2024)), ensuring all data is anonymized and pre-processed by FALCON partners in WP2-WP4. This approach guarantees a robust and ethical foundation for training our models. To defend against poisoning attacks, we employ a strategy of combining multiple ML and DL models trained on different data subsets, enhancing resilience through diversity. Additionally, we continuously monitor model performance and data distributions during and after training to detect any anomalies that might indicate data poisoning. For evasion attacks, our tool incorporates feature selection to use only those features that are less susceptible to manipulation by attackers. We also include regularization terms in the loss function to improve the model's robustness and prevent overfitting. These countermeasures collectively ensure that our ML and DL models maintain high performance and reliability, even in the face of adversarial threats.

3.4. FALCON investigative tools

In the context of FALCON's investigative tools, two key components play a crucial role in supporting Law Enforcement Agencies (LEAs) in their fight against corruption and economic crimes: the FALCON platform and the pattern identification tool.

The FALCON Platform serves as the primary user interface, providing investigators with seamless access to a range of services, including links to the tools developed under WP4 and visualization capabilities for the project's Knowledge Base. With robust backend functionalities, this platform enables efficient knowledge browsing, search capabilities via SQL and SPARQL queries, and integration with information extraction services, thereby streamlining the investigation process.

Complementing the platform is the pattern identification tool, which leverages algorithms to uncover hidden patterns within the graphs. This tool is essential for detecting suspicious behaviours by continuously analysing and updating data to reflect new information and emerging trends. However, it's important to note that the task associated with the development of the pattern identification tool began in M13 of the project. As this deliverable is submitted at M14, it covers only one month of work on this task, and therefore, detailed progress on this tool will be provided in subsequent deliverables.

3.4.1. The FALCON Platform architecture

The figure illustrates the architecture of the FALCON Visualization Dashboard and its interactions with other components in the FALCON project. The dashboard serves as a central interface for visualizing and interacting with data and services from various tools and sources.

- **Tools (Tool 1, Tool 2, Tool 3):** These represent different user interfaces (UI) from which data is gathered (WP4 tools). Each tool likely corresponds to a different module or service within the FALCON project, contributing specific types of data or insights.
- **Data Flow from\to FALCON Platform:** Data from FALCON platform, such as information about language or other parameters, is sent to the User interface tools. Then, data received from the external user interfaces are visualized in a unified view.
- **FALCON Platform:** This central component visualizes data and provides access to the integrated services. It likely offers various analytics and visualization features to help users, such as Law Enforcement Agencies (LEAs), explore, analyse, and make decisions based on the collected data.

- **Integration with Keycloak:** The dashboard is connected to a Keycloak server for identity and access management. This ensures that only authorized users can access the dashboard and its data, supporting secure and controlled access.
- **Knowledge Base:** The dashboard is also linked to the FALCON Knowledge Base, indicating that it can retrieve and display data from this repository. The Knowledge Base might store structured data, relationships, or historical information that supports investigative processes.

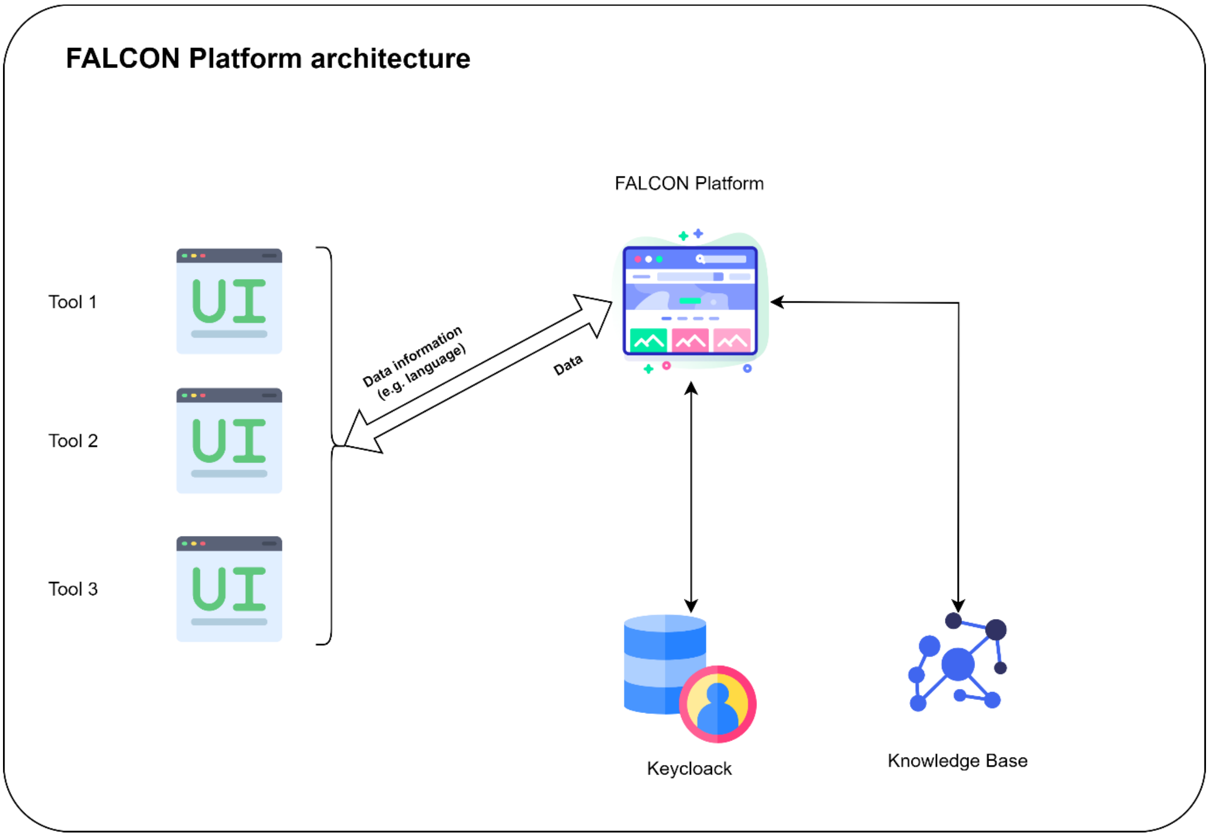

Figure 13 FALCON Platform architecture overview

3.4.2. The FALCON Platform mock-ups

In this chapter, the initial mock-ups of our application's user interface (UI) are explored, which is being developed using Next.js. The focus is on integrating Keycloak for secure authentication and authorization (as requested by Non-Functional requirements in D3.1 SEC-01 and SEC-02). The mock-up illustrates the user access flow, highlighting the seamless login and authentication process facilitated by Keycloak. This foundational design sets the stage for a robust and secure user experience within the application.

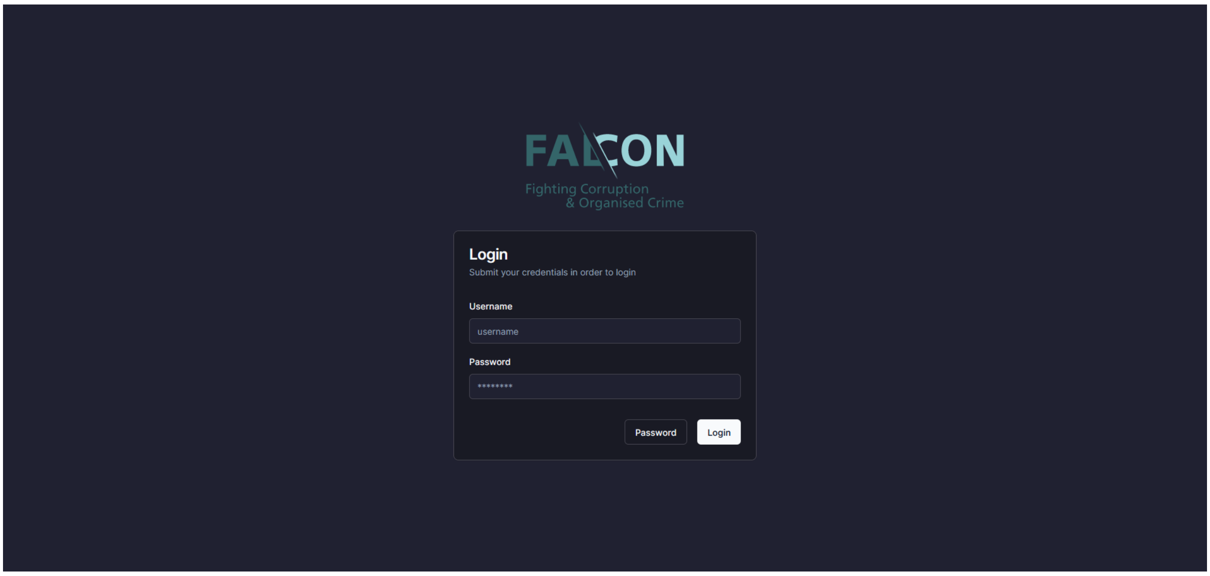

Figure 14 FALCON Platform login page.

A user who is already registered in Keycloak can easily access the platform by logging in with their existing credentials. As depicted in Figure 14, upon attempting to access the application, the user is prompted to enter their username and password on the login page. This process ensures that only authorized users can gain entry, maintaining the security and integrity of the platform. The seamless integration with Keycloak not only simplifies the login experience but also leverages a robust authentication system to protect user data. Additionally, Keycloak allows for detailed tracking of login activities, providing valuable insights and enhancing the platform's security by monitoring user access patterns.

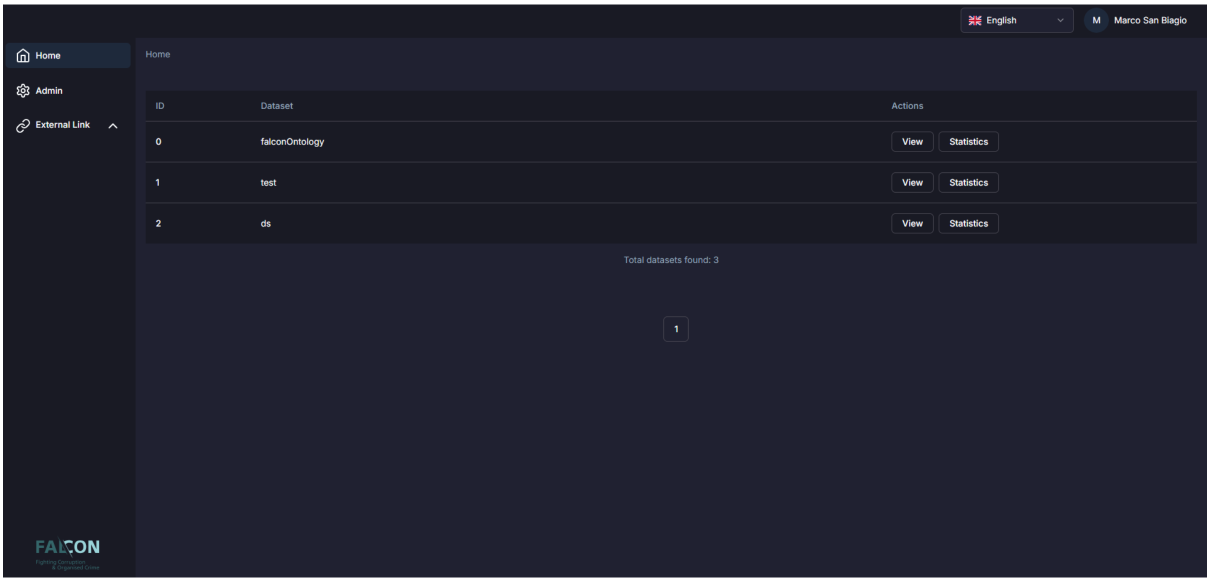

Figure 15 FALCON Platform home dashboard page

Figure 15 illustrates the mockup of the Home Dashboard, the central hub of the platform where users can access and interact with all datasets available in the knowledge base. The dashboard is designed with user flexibility in mind, offering two key functionalities: data visualization and language selection.

Users can easily navigate the available datasets, and if they wish to delve deeper into the data, they have the option to visualize the graph by selecting the corresponding dataset. This action will open a new page dedicated to graph visualization, providing a detailed view of the selected data.

Additionally, next to the visualization option, there is a **Statistics** button that allows users to view detailed statistics related to the selected dataset. These statistics are generated using SPARQL queries performed on the graph, extracting valuable insights and information to assist investigators in their analysis. This feature provides an analytical layer to the data, helping users uncover patterns and trends that may be crucial during investigations. The Statistics page is not yet available as mock-ups because we are planning to integrate an existing interactive visualization solution, such as Grafana, into the FALCON dashboard. This will allow us to leverage proven tools for data visualization and analysis, ensuring a more robust and feature-rich experience. At the top of the dashboard, users can select their preferred language from four available options: French, English, German, and Romanian. This multilingual feature fulfills the D3.1 functional requirement FUN-10, ensuring accessibility to a diverse user base.

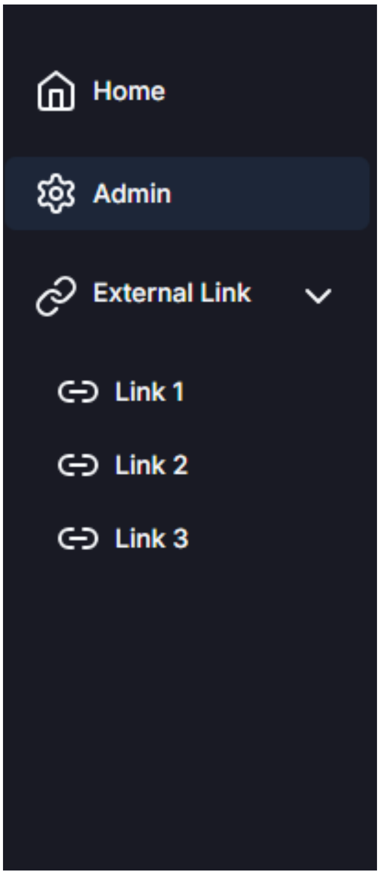

Figure 16 FALCON sidebar

On the left sidebar, Figure 16, users will find options to open settings (though these are not yet available) and to access external links. These external links, detailed in Figure 13, provide quick access to user interfaces offered by external tools, such as those developed by WP4. This feature integrates the platform with other relevant tools, enhancing its functionality and usability.

Figure 17 displays the graph visualization page. At the top, two distinct search bars are available: the first is used to filter the graph's nodes based on the data displayed in the frontend, and the second allows users to perform direct database queries using SPARQL. However, prior knowledge of SPARQL is not required to use the second search bar. Users can simply select whether they want to search for a subject, predicate, or object, and then enter a specific keyword (such as a person's name, a company name, or an object like "car").

When using the filter search, the graph automatically zooms in on the filtered node, helping users quickly locate the relevant information. The bottom part of the page shows the graph, where all the connections between nodes are visualized. Users can interact with the graph by zooming in and out to explore either the entire graph or focus on specific parts. By clicking on a specific node, users can view its detailed properties, including the name, ID, and the type of data it contains (e.g., literal, URL, etc.). This functionality allows for intuitive exploration of the relationships and attributes within the graph.

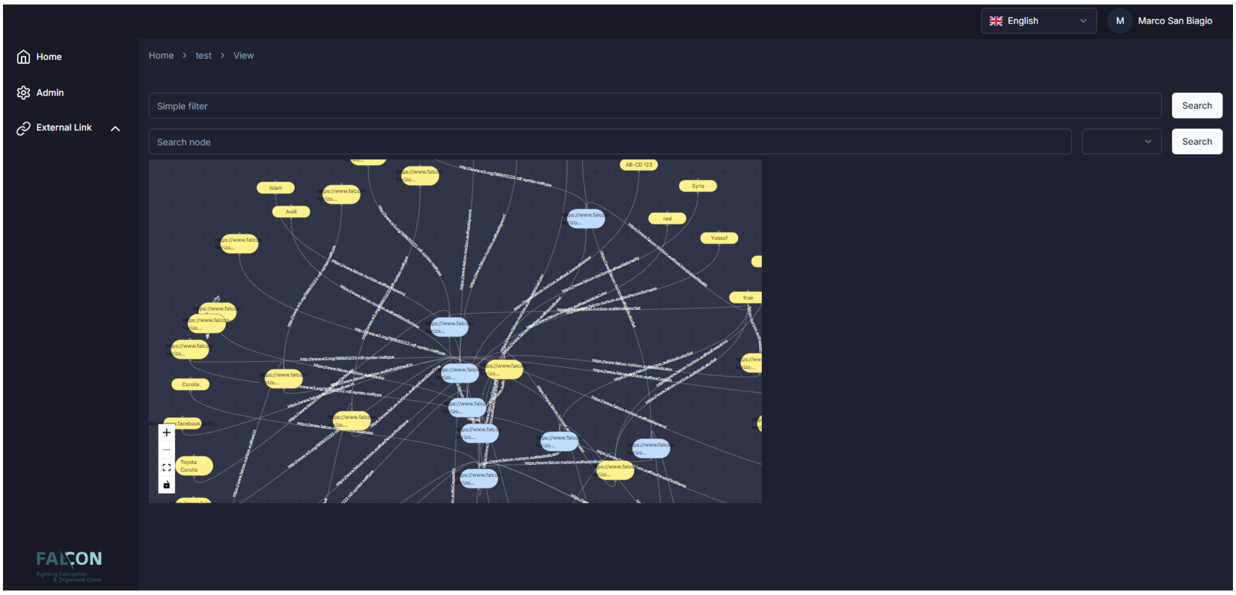

Figure 17 Graph visualization page

A few potential upgrades can further enhance the usability and functionality of the graph visualization page:

- **A legend to describe the colour of the nodes**: An essential enhancement would be the addition of a color-coded legend that explains the significance of the different node colours. Currently, the nodes are differentiated by colour to indicate whether they are subjects or objects, but without a clear legend, users may struggle to interpret this distinction. By providing a straightforward and informative legend, users would be able to quickly and accurately understand the structure of the data, greatly improving navigation and minimizing confusion.

- **A unified search bar to select between filtering or querying**: Another upgrade would be the implementation of a unified search bar, streamlining the process of filtering nodes and querying the database. Instead of having two separate search bars, users could simply choose whether they want to apply a filter on the graph or perform a more complex SPARQL-based query. This would simplify the interface, reducing confusion and enhancing user experience by offering an intuitive, all-in-one solution for both beginners and advanced users alike. An example is visualised in figure below.

Figure 18 Mock-up of the unified filter/query bar in the graph visualization page

- **Predefined searches**: A particularly valuable addition would be the inclusion of predefined searches, tailored through collaboration with our partner, the Basel Institute (BIG). These predefined searches could be crafted to align with the needs of investigators, providing them with quick access to common query types or data sets relevant to their investigations. This feature would not only save time but also ensure that investigators can leverage best practices and standardized approaches to querying the data, enhancing the efficiency and effectiveness of their investigative work.

Each of these improvements would significantly boost the overall functionality and user-friendliness of the graph visualization page, making it a more powerful tool for data exploration and investigation.

3.4.3. The FALCON Platform backend

The backend of the FALCON platform plays a crucial role in supporting the functionalities of the user interface by handling requests from the frontend and managing the data flow between the frontend and various services. Every action performed on the frontend, whether it's filtering nodes, querying the knowledge base, or visualizing data, is translated into REST API requests directed to the backend. The backend is responsible for processing these requests, interfacing with external systems, and returning the appropriate data to the frontend in a format it can use.

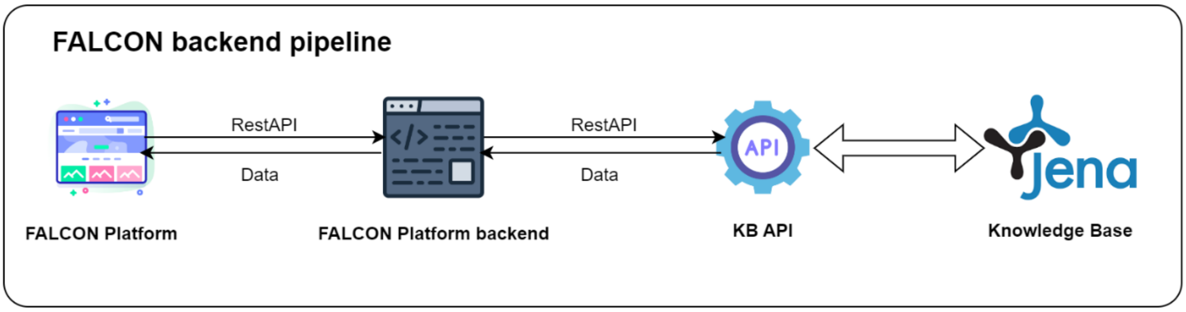

Figure 19 FALCON Backend Pipeline

The Figure 19 illustrates the **FALCON backend pipeline** and how the different components interact to handle data requests and responses within the platform.

- Starting from the left, we have the **FALCON Platform**, which represents the frontend interface that users interact with (described in paragraph 3.4.2). When a user performs an action (e.g., searching or querying data), the frontend sends a REST API request to the FALCON Platform backend.
- The FALCON Platform backend acts as an intermediary, processing the request and forwarding it as a REST API call to the KB API. The backend is responsible for tasks such as data mapping for frontend compatibility (e.g., converting data into a structure that can be visualized in frontend) and verifying user tokens to ensure access control through Keycloak.
- The KB API connects directly to the Knowledge Base, which in this case is represented by APACHE Jena (see Deliverable D4.1 for more details), a popular semantic web framework for storing and querying RDF (Resource Description Framework) data. The KB API processes the queries (often in SPARQL format) and retrieves relevant data from the knowledge base.
- Once the KB API retrieves the data, it sends it back through the pipeline to the FALCON Platform backend, which in turn formats the data and returns it to the frontend for visualization or further user interaction.

The key responsibilities of the backend include:

- **Connecting to the Knowledge Base API**: The backend interfaces with the Knowledge Base (KB) API, developed by our partner IOSB, to retrieve and query the knowledge base. Whenever a user performs a search or query through the frontend, the backend sends a request to the KB API to retrieve the relevant data. It processes queries in SPARQL or other appropriate formats to extract information from the knowledge base and return it for further processing or visualization.

- **Mapping Data for Frontend Visualization**: Once data is retrieved from the knowledge base, the backend maps it into a format that is compatible with the frontend's visualization framework. This mapping ensures that complex knowledge base relationships and entities are correctly translated into a visual representation that users can interact with seamlessly.

- **Token-Based Access Control**: The backend also manages access control for specific API endpoints by verifying user tokens through FALCON Keycloak. Each time a request is made to the backend, the token provided by the frontend is checked to ensure the user has the appropriate permissions to access the requested data or perform the desired action. This ensures that only authorized users can access sensitive information or perform specific operations within the platform.

In summary, the backend of the FALCON platform acts as the bridge between the frontend and the various services that power the application. It handles user requests, retrieves data from the knowledge base, formats the data for visualization, and ensures secure access control through token verification, creating a seamless experience for users while maintaining robust functionality and security behind the scenes.

4. Summary and conclusions

This document has provided an initial overview of the tools under development within Work Package 5 (WP5) of the FALCON project, highlighting their significance in advancing anti-corruption efforts through innovative technologies. The tools presented—namely the Advanced Corruption Risk Assessment (ACRA) tool, the Predictive Analytics Tool, and the FALCON Dashboard and Pattern Analysis Tool—are poised to substantially enhance the capabilities of Law Enforcement Agencies (LEAs) and Anti-Corruption Authorities.

The ACRA tool's ability to deliver real-time analysis and comprehensive risk assessments will empower authorities to detect and respond to corruption risks more effectively. The Predictive Analytics Tool offers valuable foresight by forecasting potential instances of single bidding, allowing pre-emptive actions to be taken at the beginning of tender processes. Meanwhile, the FALCON Dashboard and Pattern Analysis Tool will provide critical support to investigators, enabling them to visualize and analyse patterns within corruption cases and make informed decisions throughout their investigations.

This document represents Version 1 of the "Risk Assessment, Investigation, and Decision Support Toolset," serving as the foundation for the tools' development and implementation. As the project progresses, further details and enhancements will be detailed in the subsequent two documents. These future deliverables will build upon the work presented here, ensuring that the tools continue to evolve and adapt to the complex challenges of combating corruption.

In conclusion, the work outlined in this document marks a significant step forward in the fight against corruption within the European Union. By leveraging advanced technology and innovative approaches, the tools developed under WP5 are set to become invaluable assets for those tasked with upholding integrity and transparency in public processes. The continued development and refinement of these tools will be crucial in ensuring their effectiveness and impact in the ongoing battle against corruption.

5. References

Klitgaard, R. (1988). International Cooperation Against Corruption. *Finance & Development*.

Maas, H. (s.d.). *"How public organisations can use AI in anti-corruption," Hertie School*. Tratto da H. Maas, "How public organisations can use AI in anti-corrup https://www.hertie-school.org/en/debate/allcontent/detail/content/how-public-organisations-can-use-ai-in-anti-corruption.

Aa., V. (s.d.). *Digiwhist*. Tratto da https://digiwhist.eu/publications/monitoring-european-tenders-met-public-procurement-risk-assessment-software-for-authorities/

Aa., V. (s.d.). *Digiwhist*. Tratto da https://digiwhist.eu/

Aa., V. (s.d.). *European Commission*. Tratto da https://ec.europa.eu/social/main.jsp?catId=325&intPageId=3587&langId=en

Aa., V. (s.d.). *Generalitat Valenciana*. Tratto da https://participacio.gva.es/documents/162282364/164407117/Documento+explicativo-+Presentaci%25C3%25B3n+del+Sistema+de+Alertas/b1228cc0-d5bc-4121-bb63-87eb7180d262

Aa., V. (s.d.). *Hertie school*. Tratto da https://www.hertie-school.org/en/digital-governance/research/blog/detail/content/how-public-organisations-can-use-ai-in-anti-corruption-what-we-know-so-far-and-why-we-need-to-learn-more-about-it

Aa., V. (s.d.). *MOOC*. Tratto da https://www.mooc.org/

Aa., V. (s.d.). *TRACK — UNODC's central platform of tools and resources for anti-corruption knowledge*. Tratto da United Nations: https://track.unodc.org/

Aa., V. (s.d.). *Red Flags*. Tratto da https://www.redflags.eu/

Aa., V. (s.d.). *R Project*. Tratto da https://www.r-project.org/

Aa., V. (s.d.). *Transcrime*. Tratto da https://www.transcrime.it/datacros/

Aa., V. (s.d.). *Smart Team Global*. Tratto da https://www.smartteamglb.com/artificial-intelligence-imanage-ravn.html

Aa., V. (s.d.). *Universidad Complutense Madrid*. Tratto da https://www.ucm.es/otri/veripol-inteligencia-artificial-a-la-caza-de-denuncias-falsas

Aa., V. (s.d.). *Wikipedia.org*. Tratto da https://en.wikipedia.org/wiki/Document_management_system

Aa., V. (s.d.). *ANAC*. Tratto da https://dati.anticorruzione.it/opendata/ocds_en

Sanz, I. P., & Félix, J. L.-I. (2019). Social Indicators Research: An International and Interdisciplinary Journal for Quality-of-Life Measurement. In *Predicting Public Corruption with Neural Networks: An Analysis of Spanish Provinces.* Springer.

Carneiro, D., Veloso, P., Ventura, A., Palumbo, G., & Costa, J. (2020). Network Analysis for Fraud Detection in Portuguese Public Procurement. *Lecture Notes in Computer Scienc*. Tratto da https://link.springer.com/chapter/10.1007/978-3-030-62365-4_37.

Dhurandhar, A., Graves, B., Ravi, R., Maniachari, G., & Ettl, M. (2015). Big data system for analyzing risky procurement entities. *KDD*.

Bai, J., & Qiu, T. (2023). Automatic Procurement Fraud Detection with Machine Learning. *arXiv* .

Aran, M. I., Carabetta, J., Wen, J., Julià-Verdaguer, A., Rosado, P., Sidgwick, J., & Ghani, R. (2020). Reducing corruption in public procurement using machine learning. *Zenodo*.

Gallego, J., Rivero, G., & Martínez, J. (2018). *Preventing rather than Punishing: An Early Warning Model of Malfeasance in Public Procurement.* Universidad del Rosario .

Lima, M., Silva, R., Mendes, F. L., Carvalho, L. R., Araujo, A., & Vidal, F. d. (2020). *Inferring about fraudulent collusion risk on Brazilian public works contracts in official texts using a Bi-LSTM approach.* Findings of the Association for Computational Linguistics: EMNLP.

García Rodríguez, M. J., Rodríguez-Montequín, V., Ballesteros-Pérez, P., Love, P. E., & Signor, R. (2022). Collusion detection in public procurement auctions with machine learning algorithms. *Automation in Construction*.

Decarolis, F., & Giorgiantonio, C. (2022). Corruption red flags in public procurement: new evidence from Italian calls for tenders. *EPJ Data Science*(16).

Vv., A. (2024). *Rozes*. Tratto da https://www.rozes.ai/

Ferwerda, J., Deleanu, I., & Unger, B. (2017). Corruption in public procurement: finding the right indicators. *European journal on criminal policy and research, 23*, 245-267.

Goryunova, N., Baklanov, A., & Ianovski, E. (2021). Detecting corruption in single-bidder auctions via positive-unlabelled learning. *International Conference on Mathematical Optimization Theory and Operations Research* (p. 316-326). Cham: Springer International Publishing.

Lyra, M. S., Damásio, B., Pinheiro, F. L., & Bacao, F. (2022). Fraud, corruption, and collusion in public procurement activities, a systematic literature review on data-driven methods. *Applied Network Science, 7*(83).

Opentender. (s.d.). *Opentender.* Tratto il giorno 2024 da https://opentender.eu/

Project, G. (s.d.). *"The GDELT Project"*. Tratto il giorno 2024 da https://www.gdeltproject.org/

Benítez-Martínez, F. L., Romero-Frías, E., & Hurtado-Torres, M. V. (Jan 2023). Neural blockchain technology for a new anticorruption token: towards a novel governance model. *Journal of Information Technology & Politics, 20*(1), 1-18.

Weingärtner, T., Batista, D., Köchli, S., & Voutat, G. (2021). Prototyping a Smart Contract Based Public Procurement to Fight Corruption. *Computers, 10*(7).

Mazza, E., Cappellari, R., & Renzi, D. (2022). A Blockchain-Based Approach to Enhancing Transparency and Integrity in Public Procurement. *Journal of Public Procurement and Policy, 25*(3), 245-259.

Caruso, S., Bruccoleri, M., Pietrosi, A., & Scaccianoce, A. (May 2023). Artificial intelligence to counteract 'KPI overload' in business process monitoring: the case of anti-corruption in public organizations. *Business Process Management Journal, 29*.

Dawson, L., & Rahman, B. (2022). Artificial Intelligence and the Future of Anti-Corruption: Exploring the Role of Predictive Analytics in Law Enforcement. *International Journal of Law and Technology, 33*, 182-202.

Mungiu-Pippidi, S. (2021). The Good, the Bad, and the Ugly: Controlling Corruption in the European Union. *Journal of Democracy, 27*(1), 50-64.

Wacker, J., Ferreira, R. P., & Ladeira, M. (2018). Detecting Fake Suppliers using Deep Image Features. *7th Brazilian Conference on Intelligent Systems (BRACIS)* , 224-229.

Gallab, M., Bouloiz, H., Alaoui, Y. L., & Tkiouat, M. (2019). Risk Assessment of Maintenance Activities using Fuzzy Logic. *Procedia Computer Science, 148*, 226-235.

Aa., V. (2024). *Kleptotrace*. Tratto da https://transcrime.it/kleptotrace/

Aa., V. (2024). *Transparency International UK*. Tratto da https://www.transparency.org.uk/concerns-corruption-all-time-high-uk-falls-its-lowest-ever-score-global-corruption-perceptions-index

Aa., V. (s.d.). *Twenty years of UNCAC: uniting the world against corruption*. Tratto da https://www.corruptionwatch.org.za/twenty-years-of-uncac-uniting-the-world-against-corruption/

Aa., V. (2024). *Opentender*. Tratto da https://opentender.eu/