



Funded by
the European Union

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



Deliverable D4.2

Title: Corruption data acquisition and analysis toolset (R1.0)

Dissemination Level: PU - Public

Nature of the Deliverable: R

Date: 31/08/2024

Work Package: WP4 - Corruption data acquisition and
analysis tools

Editors: CERTH

Reviewers: ICCS, ENG, UCSC, C&T

Contributors: UPV, GTI, UCSC, C&T, BPTI, SPH, VICOM

Abstract: This document presents the first version of the "Corruption Data Acquisition and Analysis Toolset (R1.0)," detailing the initial FALCON tools designed for extracting indicators, detecting anomalies, and analysing trends. It outlines the work completed in Work Package 4 (WP4) through various tasks, emphasizing the alignment of these tools with the project's objectives as specified in the Description of Action (DoA). The report includes a description of data management tools, tailored indicator extraction tools for specific corruption-related use cases, and a comprehensive overview of methodologies for anomaly detection and trend analysis.

Executive Summary

This document is the first version of the “Corruption data acquisition and analysis toolset (R1.0)” the report on the 1st release of FALCON tools for extraction of indicators from underlying data, anomaly detection, and trends analysis. The objective of this document is to provide a description of the tools developed in WP4 and present the work performed in Tasks T4.1, T4.3, T4.4 and T4.5 until month 12 of the project.

The tools outlined in this document have been meticulously developed to address the specific needs of the FALCON project, as articulated in the Description of Action (DoA). These tools are informed by a comprehensive analysis of all project deliverables produced to date, with particular emphasis on Deliverables D3.1 (Use cases and requirements), D3.2 (FALCON framework architecture) and D2.2 (Comprehensive list and definitions of corruption risk indicators). By aligning our tools with the project's objectives and requirements, we aim to enhance efficiency, facilitate collaboration, and ensure the successful implementation of the FALCON's goals.

Initially, the tools for ensuring and managing access to datasets within the overall FALCON framework architecture are described. These tools include all the required adapters for the FALCON heterogeneous data sources, a dataflow manager that will take care of data injection and data modelling that covers FALCON's corruption crime description needs. Furthermore, the dataflow management is described as well as the FALCON middleware. Finally, a short description of the datasets used for data collection is provided.

This document also reports on the work conducted in extracting indicators. A suite of tailored indicator extraction tools has been developed to meet the specific needs of end-users and data availability. These tools focus on border corruption indicators and car recognition (UC3), public procurement risks (UC1 and UC4), cryptocurrency and the analysis of social data networks encompassing all FALCON use cases.

Finally, a comprehensive State of the Art (SoA) is presented regarding the methodologies for anomaly detection and trend analysis that will be integrated into the upcoming versions of the FALCON tools. This detailed overview encompasses various innovative techniques, including machine learning algorithms, statistical analysis methods, and data mining approaches. The focus will be on enhancing the accuracy and efficiency of detecting anomalies in large datasets, as well as identifying emerging trends that can provide valuable insights for decision-making processes.