

**U4 HELPDESK ANSWER 2025: 34**

# **The use of big data by anti-corruption authorities**

**Iva Parvanova**

Reviewed by

**Caitlin Maslen, Roberto Kukutschka and David Eristavi (TI)**

**Sofie Schütte (U4)**

Anti-corruption authorities (ACAs) are increasingly using big data to prevent, detect and investigate corruption, particularly through the analysis of procurement, asset declaration, company ownership and financial intelligence datasets. These practices are most advanced in countries with mature digital registers and interoperable systems and where ACAs can combine in-house analytics with structured cooperation with tax, procurement, financial and audit bodies. For development professionals, this underscores the importance of supporting not only legal reforms and IT platforms but also data quality, interoperability, analytical skills and safeguards for privacy and due process.

U4 Helpdesk Answers are tailor-made research briefings compiled in ten working days. The U4 Helpdesk is a free research service run in collaboration with Transparency International.

[helpdesk@u4.no](mailto:helpdesk@u4.no)

#### Related U4 reading

[Anti-corruption agencies and public engagement \(2025\)](#)

[The promise and perils of data for anti-corruption efforts in international development work](#)

[Bespoke monitoring and evaluation of anti-corruption agencies \(2017\)](#)

#### How to cite

Parvanova, I. 2025. The use of big data by anti-corruption authorities. Bergen: Transparency International and U4 Anti-Corruption Resource Centre, Chr. Michelsen Institute (U4 Helpdesk Answer 2025:34)

#### Published

22 December 2025

#### Keywords

Anti-corruption authorities - big-data analytics - data governance

## Query

Which types of big data are anti-corruption authorities (ACAs) using?

Which forms of corruption do these data applications address, and how is this work organised institutionally?

### Main points

- Anti-corruption authorities (ACAs) are increasingly using big data proactively, not just reactively. E-procurement systems, asset and interest declaration platforms and beneficial ownership (BO) registers now generate large volumes of machine-readable information. ACAs can use these to identify risks, prioritise cases and demonstrate results, but their ability to do so varies greatly across country contexts.
- ACAs typically rely on a small set of core data families used across different functions. The main sources are: (i) public procurement and contracting data; (ii) company and BO registers; (iii) income, interest and asset declarations; and (iv) financial intelligence and related datasets. These support corruption prevention (risk mapping, red-flagging, ex ante conflict of interest screening), investigations (network and pattern analysis) and prosecution (using evidence to build and open cases).
- Concrete country experiences show the promise of big data approaches. Examples include AI-supported procurement analytics, integrity screening in procurement based on asset-declaration data, large-scale e-declaration risk-scoring and BO-based network analysis. These demonstrate how linking datasets can prevent conflicts of interest, guide investigations and support asset recovery.
- Institutional arrangements for cross-government data sharing largely determine

what ACAs can do with data. The most relevant datasets are usually held by other institutions (tax, customs, registries, FIUs, courts, procurement bodies). Effective use of big data depends on legal gateways, technical interoperability and governance arrangements that enable secure, purpose-limited sharing. Emerging practices include memoranda of understanding, joint analytical units and shared data hubs, but these must navigate legal, technical, organisational and political obstacles.

- In practice, the kind of big data ACAs would need to fully exploit advanced analytics is still rare, even in relatively digitalized contexts, such as the EU. The main constraint is often the broader integrity-data ecosystem. While procurement data may be available in bulk, additional linkable datasets needed for integrity analytics are frequently missing or unusable.

# Contents

<b>Background</b>	<b>5</b>
Anti-corruption authorities and big data	5
What is “big data” in anti-corruption?	6
<b>Types of data used by ACAs</b>	<b>8</b>
Public procurement and contracting data	8
Company and beneficial ownership data	10
Officials’ income, interest and asset declarations	11
Suspicious transactions reports	13
<b>Institutional determinants of ACAs’ use of big data</b>	<b>15</b>
Types of institutional set-ups	15
Cross-institutional data sharing	16
Risks associated with using big data	19
<b>Conclusion</b>	<b>20</b>
<b>References</b>	<b>21</b>

# Background

## Anti-corruption authorities and big data

Anti-corruption authorities (ACAs) are public institutions mandated to prevent and counter corruption. They have proliferated since the 1990s, often in response to the United Nations Convention against Corruption (UNCAC), which under articles 6 and 36 calls on states to establish bodies to coordinate corruption prevention and law enforcement, and to ensure they have sufficient independence to do so (Maslen 2025). Today, ACAs operate alongside other institutions with relevant mandates – such as police units, supreme audit institutions, inspectorates general and the judiciary – within broader national integrity systems. Their structures vary, but most combine some mix of prevention, education and awareness raising, investigation and, in some cases, prosecution (United Nations Office on Drugs and Crime 2020a; Schütte & David-Barrett 2025):

- Prevention: ACAs lead efforts to develop, implement, monitor and coordinate anti-corruption policies.
- Education and awareness raising: ACAs promote anti-corruption efforts within the government bureaucracy as well as including activities with the private sector and/or the public.
- Investigation: ACAs investigate allegations of corruption, whether on their own initiative or in response to a complaint.
- Prosecution: some prosecutorial services have established specialised anti-corruption units and have seconded prosecutors directly to an ACA. In other cases, ACAs are endowed with the power to prosecute.

The question of how ACAs work with big data can be linked to ongoing debates around the effectiveness of these authorities (on the topic of ACAs' effectiveness see Schütte et al. 2023). ACAs are increasingly expected to deliver impact (World Bank 2020), and big data can be useful in making this measurable. While these authorities rarely generate large datasets, their growing use of such data can help improve issues of the evaluability of their effectiveness (Johnsøn et al. 2011: 17 - 21). Moreover, data-driven approaches would make ACAs better positioned to allocate scarce investigative and prevention resources in a cost-effective manner. Against this backdrop, big-data initiatives in ACAs can be understood as one way of implementing the Jakarta

Statement on the Principles of ACAs<sup>1</sup> (United Nations Office on Drugs and Crime 2012). Yet not all ACAs are equally positioned to engage with big data. While some operate in highly digitalised environments that meet the necessary conditions, such as the availability of centralised procurement portals and registry databases, others work with fragmented records or largely paper-based systems.

## What is “big data” in anti-corruption?

The information environment in which ACAs operate has changed profoundly. Public administrations globally generate large volumes of machine-readable data through e-procurement platforms, electronic asset and interest declaration systems, and beneficial ownership registers, among other platforms (Adam & Fazekas 2021). In technical debates, big data is often defined by the “three Vs”: volume, velocity and variety (Sagiroglu and Sinanc 2013). In the field of anti-corruption, these translate to:

- Volume: large quantities of administrative and transactional records, such as millions of procurement procedures, asset declarations, tax filings or suspicious transaction reports
- Velocity: the speed at which new records are generated; for instance, daily updates to e-procurement portals, near-real-time payment systems or continuous reporting of financial transactions to financial intelligence units (FIUs)
- Variety: the mix of structured tables (contracts, registry entries, payment records), semi-structured formats (XML, PDFs with embedded tables) and unstructured text (complaints, legal justifications, media content), as well as geospatial data (satellite imagery, GPS-tagged photos)

Beyond ACAs themselves, big data has been used in a range of anti-corruption applications that illustrate its potential. Procurement contract datasets have underpinned red-flag and risk-scoring systems that detect patterns such as single bidding, collusion and overpricing at scale used by academic researchers (e.g., Fazekas and Kocsis 2020) and international organisations (e.g., ProAct n.d.). Geospatial and satellite data have been used by NGOs to spot illegal logging, mining or “ghost” infrastructure projects that suggest corruption in licensing or public

---

<sup>1</sup> For example, large administrative datasets to inform risk-based planning helps ACAs give a concrete effect to the mandate principle of the Jakarta statement and to the principle on adequate and reliable resources, which stresses timely, planned and adequate resourcing for the agency’s operations (United Nations Office on Drugs and Crime 2012: 2). Additionally, establishing structured data-sharing arrangements can be viewed as a way to operationalise the collaboration principle (United Nations Office on Drugs and Crime 2012: 2), according to which ACAs shall foster good collaborative relations with intersectoral stakeholders.

investment.<sup>2</sup> Beneficial ownership and company data support the identification of opaque corporate networks and politically exposed firms, while text data from disclosures and court documents can be mined for hidden links and conflict of interest risks.<sup>3</sup> Sanctions lists and debarment databases add another layer, helping to target high-risk actors and markets.<sup>4</sup> Together, these applications exemplify a broader shift from responses to individual allegations towards systematic, data-driven risk detection and prioritisation (OECD 2025).

It is important to emphasise that, in practice, the kind of “big data” ACAs would need to fully exploit advanced analytics is still relatively rare. Even in the EU, where digitalisation is comparatively advanced, a recent assessment of data usability shows that while all member states now publish bulk, machine-readable procurement data, far fewer provide the additional, linkable layers that integrity analytics require – such as usable beneficial ownership, political finance, complaint and media-ownership datasets (Longobucco & Ferwerda 2025: 11). The constraint is therefore often the wider integrity-data ecosystem rather than ACAs simply lagging behind.

This Helpdesk Answer focuses on three interrelated aspects of ACAs’ practices regarding big data. First, it lays out the types of data ACAs typically draw on in their work, according to the type of activity the authority engages in. Notably, the following sections examine how ACAs are using data, whether through basic matching and statistics or more sophisticated modelling. Additionally, in some cases, advanced analytic techniques, such as artificial intelligence (AI) can be employed; however, to manage scope, this analysis does not discuss AI techniques in depth. Second, the analysis traces how specific forms of corruption, such as procurement manipulation, conflicts of interest, illicit enrichment, money laundering and systemic favoritism, are tackled through these data across ACA mandates. Table 1 then synthesizes these linkages by mapping data sources, analytical techniques, ACA functions and country examples in one place. Lastly, the paper examines institutional drivers and deterrents behind data-sharing institutional arrangements. These include different organizational set-ups, legal and technical enablers, barriers and risks that determine how ACAs engage with big data.

---

<sup>2</sup> For example, a civil society research organisation in Brazil, Imazon, combined satellite imagery with large-scale logging-authorisation data to detect illegal logging in the Amazon (Imazon 2023).

<sup>3</sup> Law enforcement agencies in the UK report using the British register of beneficial ownership to identify and investigate individuals and networks involved in money laundering (Department for Business, Energy & Industrial Strategy 2019: 25-26).

<sup>4</sup> The Global Sanctions Database (GSDB), compiled by researchers, contains case-level information on sanctions imposed globally and is used in quantitative research to study the wider impacts of sanctions, including their role in anti-corruption efforts (Felbermayr et al. 2020).

# Types of data used by ACAs

A review of the academic and grey literature reveals that ACAs across a variety of settings draw on a similar set of core data families: public procurement and contracting records, company and beneficial ownership information, income, interest and asset declarations, and financial intelligence and related datasets.<sup>5</sup> These sources are used across the full range of ACA functions – prevention, education and awareness raising, investigation and prosecution – although the emphasis and depth of use differs by country, mandate and data availability.

## Public procurement and contracting data

E-procurement portals and central contract registers usually record information on public tenders, including the number of bids, awards, contract values, timeframes and procedure types (Fazekas et al. 2016). In the area of corruption prevention, these datasets are used to map procurement markets, identify structural risks and build red-flag indicators.

Table 1 presents the variety of corruption types which can be addressed through the use of procurement data in further detail. ACAs can use these portals to monitor single-bid rates, the concentration of awards among a small group of suppliers or the extensive use of non-competitive procedures and framework contracts. Italy's National Anti-Corruption Authority (Italian: *L'Autorità Nazionale Anticorruzione*, ANAC) maintains a national database of public contracts which consolidates information from contracting authorities across the country and underpins regular analyses of competition levels and modification practices (Autorità Nazionale Anticorruzione n.d.). Brazil's Comptroller General of the Union (Portuguese: *Controladoria-Geral da União*, CGU) has gone one step further by creating an AI supported application called ALICE, or Analysis of Biddings and Call for Bids (Odilla 2023). This application mines procurement notices and contracts from federal portals and the official gazette. It also screens them for combinations of risk indicators and feeds dashboards and alerts to auditors before awards are finalised. The system has demonstrated a positive impact on procurement integrity at state and

---

<sup>5</sup> In reviewing this literature, the author adopted an inductive approach, starting from a list of datasets commonly used by other stakeholders (e.g. procurement, beneficial ownership, political finance, lobby, tax and land registers) and then reviewing the literature to examine whether, and how, these same data types are used by ACAs. These five types of data were identified as those most used by ACAs. However, it should be noted that this list is not exhaustive as evidence indicates ACAs use other datasets, for example, political finance, lobby, tax and land registers.

federal levels. In the first year after ALICE was introduced, 100,000 notices were analysed, of which eight bids were revoked totalling approximately R\$3.2 billion (US\$592 million) (OECD 2022a: 25).

The same procurement data can be repurposed for education and awareness raising. For example, ANAC publishes dashboards and interactive maps that display contract-level information (Autorità Nazionale Anticorruzione n.d.). These have been used in outreach and in the training of media stakeholders, such as the Lazio Order of Journalists, on how to interact with the database and relevant indicators (Autorità Nazionale Anticorruzione 2023). ACAs can also collaborate with and use data generated by civic-tech platforms in their public engagement.

When it comes to the investigative function of ACAs, procurement datasets that initially served for broad risk screening can also serve as sources for evidence. Investigators use them to reconstruct bidding histories of suspect firms, compare prices and specifications across similar tenders and examine the timing and nature of amendments and payments. In Lithuania, the Special Investigation Service (Lithuanian: *Specialiųjų tyrimų tarnyba*, STT) has developed, with support from the European anti-fraud office, an analytical tool that integrates procurement and EU funds data with criminal intelligence information, generating risk profiles for projects and contracting bodies that guide targeted checks (European Commission Anti-Fraud Knowledge Centre 2021). These tools help investigators focus their efforts on specific tenders, markets or suppliers with high-risk profiles.

In the prosecution phase, procurement records form the backbone of many corruption, fraud and bid-rigging cases. Tender documents, bids, evaluation reports, award decisions and contract amendments can provide the documentary trail needed to prove elements such as favouritism, manipulation of criteria, collusive patterns or unjustified price increases. Big data applications largely operate upstream, by highlighting which procurement processes merit close attention; in court, it is the underlying records rather than the overarching analytical results themselves that are introduced as evidence. In Ukraine, National Anti-Corruption Bureau of Ukraine (Ukrainian: *Національне антикорупційне бюро України*, NABU) has opened prosecutions on the basis of red flags raised via DOZORRO, a civic-tech platform built on the national Prozorro e-procurement data. In one case pertaining to the construction of a kindergarten, a local activist's complaint on DOZORRO triggered scrutiny by a civil society organisation (CSO) whose findings were later confirmed by the state audit service. This eventually led to a NABU investigation that uncovered an embezzlement scheme worth over UAH27 million (around US\$635,000) and led to charges before the High Anti-Corruption Court (Transparency International Ukraine 2022).

## Company and beneficial ownership data

Company registers and beneficial ownership (BO) databases help ACAs identify the true ownership of companies as well as detect embezzlement and laundering of funds through shell companies, among other corruption types. When linked to procurement, licensing or subsidy data, these records allow analysts to map networks of influence and control (Arista et al. 2024). BO datasets include different information across various jurisdictions. In Europe, minimum requirements include the beneficial owner's name, date of birth, nationality and residence status as well as the nature and extent of the beneficial interest (European Parliament and Council of the European Union 2015: 97).

In practice, ACAs' access to BO data depends on national legal frameworks (Open Ownership 2024). BO registers are often formally administered by corporate registries, tax authorities or central banks. ACAs may be granted direct access, or they may need to request extracts for specific cases. In Europe, recent court rulings have curtailed unrestricted public access to BO data on data-protection grounds, while access for competent authorities has been maintained. This category includes anti-money laundering actors such as FIUs (Financial Action Task Force n.d.). As a result, automatic access to BO data for anti-corruption authorities (ACAs) depends on whether they are legally designated as competent authorities under national arrangements and on the scope of their mandates. At the same time, it has been argued that extending access beyond AML prevention and investigation would strengthen ACAs' effectiveness (Transparency International 2023a). When ACAs use BO data, they may encounter technical and quality problems, such as missing or inaccurate reports. These limitations further inhibit the usefulness of BO data in some settings. Addressing these issues requires sustained efforts by registry authorities and often goes beyond the remit of ACAs (United Nations Department of Economic and Social Affairs 2024:8).

Furthermore, country practices illustrate a trade-off between data protection and comprehensiveness. For example, the UK's People with Significant Control (PSC) register records only personal details and discloses ownership and voting rights in broad percentage bands, which limits granular analysis (Department for Business and Trade 2025). On the other hand, Ukraine's unified state register includes detailed identifiers, such as passport and taxpayer numbers, improving traceability but raising stronger privacy risks (World Bank 2017: 7).

ACAs focusing on prevention rely on BO data generated by projects such as DATACROS. The first iteration of the consortium was developed by the European Commission in collaboration with the French Anti-Corruption Agency (French: *Agence française anticorruption*, AFA) among other partners (DATACROS I n.d.).

Currently, it is led by Transcrime<sup>6</sup> and 22 partners, including Italy's ANAC (DATACROS III n.d.). It uses firm-level indicators and network analytics to identify anomalous ownership structures associated with heightened risks of collusion, corruption, money laundering and other financial crimes (Bosisio et al. 2021). These cover over 400 million firms across more than 200 countries (DATACROS III n.d.).

When cases reach the prosecution and (where relevant) the asset recovery stage, some ACAs use extracts from company and BO databases as evidentiary support for allegations of collusion and money laundering. They can help prosecutors demonstrate relationships, such as firms that have presented as competitors but share an owner, or trace chains of entities connecting public decisions to private gain. In Malaysia, the Malaysian Anti-Corruption Commission (Malay: *Suruhanjaya Pencegahan Rasuah Malaysia*, MACC) worked with the police and the central bank in the 1MDB case to trace the beneficial owners of companies and accounts linked to the embezzlement scheme, using beneficial ownership information to help identify key individuals and corporate vehicles involved. This analysis supported the successful prosecution of former prime minister Najib Razak and contributed to the recovery of approximately MYR1.2 billion (around US\$291million) in misappropriated assets (U4 Anti-Corruption Resource Centre 2023).

## Officials' income, interest and asset declarations

Income, interest and asset declaration (AID) systems are now a central part of ACAs' use of big data. Electronic filing, structured databases and systematic cross-checks against tax, land, corporate and other registers, have turned declarations into a multi-purpose tool for preventing conflicts of interest, detecting unjustified enrichment and supporting asset recovery (Burdescu et al. 2009). In this setting, AID systems become a core component of ACAs' data architectures, rather than a passive compliance step (World Bank 2021a; World Bank 2023a).

Table 1 links AID systems to all relevant corruption types and they are useful for different types of ACAs. For example, in Indonesia, the cross-functional Corruption Eradication Commission (Indonesian: *Komisi Pemberantasan Korupsi*, KPK) is responsible for the LHKPN wealth-reporting system for certain categories of officials. Initially paper-based, KPK progressively built an electronic filing platform (e-LHKPN) and a back-end data-warehouse and reporting system. This allows analysts to process large volumes of declaration data, monitor submission compliance rates across institutions and regions, and generate reports on total assets and changes in declared wealth (UNODC 2019). Recent work notes the growing use of data-analytic

---

<sup>6</sup> Transcrime is a joint research centre of three Italian universities working on innovation and crime.

techniques on LHKPN data and highlights that the KPK uses in both preventive screening (for example, checking candidates for sensitive positions) and investigative case-building (Putra 2024; UNODC 2020b).

Prevention-focused ACAs may use AID systems in conjunction with other big-data sources, such as procurement systems, to trace conflicts of interest. For example, Romania's *Agenția Națională de Integritate* (National Integrity Agency, ANI) operates the PREVENT system which is built on data sharing agreements allowing for cross-referencing information across institutions (European Commission Anti-Fraud Knowledge Centre 2020; Transparency International Romania 2022). The primary sources of information are integrity forms submitted to the Public Procurement Electronic Service. These are then linked to data held in the Department for Population Records and Database Management (Romanian: *Direcția Generală pentru Evidența Persoanelor și Administrarea Bazelor de Date*, DGEP) and the National Trade Register Office (Romanian: *Oficiul National al Registrului Comertului*). On the basis of the relevant legal provisions, the system draws directly on these state registers rather than relying on ad hoc data requests, which enables more systematic detection of potential conflicts of interest (Parliament of Romania 2016: 4)

When ANI identifies and signals possible conflicts of interest, such as family links between decision-makers and bidders through this system, contracting authorities are required to address the risk. This can include replacing a conflicted official or excluding a bidder before the contract is signed (European Commission Anti-Fraud Knowledge Centre 2020). The system significantly reduced undisclosed conflicts of interest in procurement, illustrating how linking declaration data to other large administrative datasets enables real-time, *ex ante* integrity controls in public contracting (European Commission Anti-Fraud Knowledge Centre 2020).

In Ukraine, *Національне агентство з питань запобігання корупції* (the National Agency on Corruption Prevention, NACP) runs a large-scale e-declaration regime collecting a large number of electronic declarations annually. The central register is connected to state databases, including a property register and the land cadastre, and uses automated “logical and arithmetic controls” and cross-checks with external registries to assign risk ratings to each declaration. While the default verification is automated, NACP prioritises full verification on high-risk cases while lower-risk declarations are cleared through automated checks (World Bank 2021a; NACP 2023).

Systems like PREVENT underscore the importance of technical infrastructure, governance arrangements and human capacity. For AID data to be used in an effective and relevant manner, interoperable IT architectures and common identifiers are necessary to link multiple registers. Legal bases for data sharing are necessary

since ACAs tend to be users of AID and other complementary data which is often generated by tax and procurement bodies (World Bank 2023a; World Bank 2021a).

## Suspicious transactions reports

The use of suspicious transactions reports (STRs) by ACAs is not well documented, despite the fact that these cannot be only a strategic input but also an operational trigger that can be combined with other datasets. In many jurisdictions, suspicious transaction reports (STRs) are first collected and analysed by a financial intelligence unit (FIU), which then disseminates relevant reports or intelligence packages to bodies with a corruption mandate. In Latvia, *Korupcijas novēršanas un apkarošanas birojs* (the Corruption Prevention and Combating Bureau, KNAB) is authorised to conduct money laundering investigations. The ACA receives financial intelligence related to public officials and politically exposed persons, which it cross-checks against asset declarations, political finance reports and procurement information to identify unexplained wealth or possible kickback schemes in public contracting (Moneyval 2018: 44). However, it has been documented that the actions of KNAB are not fully consistent with Latvia's risk profile as it under-investigates ML cases domestically, despite the significant risk they pose (Moneyval 2018: 39). Ultimately, it often delegates cases to other authorities which are more specialised in anti-money laundering (AML) (Moneyval 2018: 56).

Table 1: Forms of corruption that ACAs can address with different data sources

Type of corruption / problem	Main data sources used	Typical analytical techniques	ACA functions most affected	Illustrative ACA examples
Procurement manipulation and collusion	E-procurement and contract registers; supplier master data; complaint and audit logs	Red-flag indicators; outlier detection; bid-network analysis; text mining of tender notices	Prevention (e.g. risk mapping and case selection)	Brazil - CGU: ALICE analyses of procurement notices, contracts and price registrations to flag irregularities and suspend risky purchases pending audit (Odilla 2023)
Conflicts of interest and undeclared interests	Asset and interest declarations; company and beneficial ownership registries; HR and procurement data; lobbying / meetings registers where available	Entity resolution; relationship mapping; ex ante conflict screening; rule-based flags	Prevention (e.g. integrity management, administrative enforcement)	Romania - ANI: PREVENT cross-checks procurement data with civil status and company registers to detect family-based conflicts of interest before contract award (European Commission Anti-Fraud Knowledge Centre 2020).
Illicit enrichment and unjustified wealth	Asset declarations; tax records; property and vehicle registries; limited financial data (via FIUs)	Risk-scoring of declarations; wealth-income comparisons; cross-register matching	Investigation (e.g. support to asset recovery)	Armenia - CPC: verification department uses asset and interest declarations, tax and registry data with risk-based selection for in-depth checks (OECD 2022b; World Bank 2023b)
Embezzlement, bribery and laundering of proceeds	STRs and other FIU data; procurement and budget data; company / BO records	Transaction network analysis; anomaly detection in payment flows; link analysis	Investigation (e.g. financial case building, asset tracing)	Latvia – KNAB: uses STRs with other data to compile investigations into suspected money laundering (Moneyval 2018)
Systemic favouritism and policy capture	Aggregated procurement data; lobbying / interest registers; political finance data; complaints and audit recommendations	Sectoral risk indices; concentration measures; trend and cluster analysis	Awareness raising (e.g. policy advice, public reporting) and prevention (e.g. strategic analysis)	Italy - ANAC: uses composite procurement risk indicators and contextual data to identify high-risk sectors and contracting authorities (ANAC 2021)

# Institutional determinants of ACAs' use of big data

## Types of institutional set-ups

When considering how ACAs use “big data”, it is also important to understand the institutional set-up and environment within which this takes place. The ways in which ACAs operate can be grouped into three (non-mutually exclusive) categories: in-house, collaborative and outsourcing<sup>7</sup>. As already discussed, ACAs rarely generate the big data they use as this underlying information comes from large administrative datasets. However, in the in-house set-up, ACAs lead and control the analysis. For example, Italy’s ANAC uses procurement awards notices created by the state to generate a public procurement database and associated dashboards. ANAC creates, hosts and analyses these secondary datasets which allow the calculation of corruption-risk indicators at detailed territorial and sectoral levels (Autorità Nazionale Anticorruzione 2021).

On the other hand, in a collaborative model, ACAs rely on specialist institutions for sensitive or complex datasets (such as bank transactions, tax records, customs data and FIU reports) and work through joint intelligence centres. These inter-agency centres collect, store and analyse information and coordinate the sharing of it between stakeholders, bringing together tax administrations, FIUs, ACAs and prosecutors (Gunn and Scott 2018: 88). ACAs build analytics in collaboration with governmental agencies, which reduces duplication and supports benchmarking. Capacity building is often built into this collaborative model. For example, Lithuania’s big-data analytics tool for detecting corruption and fraud risks was developed with OLAF funding, which covered both the creation of the tool and specialised training for ACA staff in how to use big-data analytics (European Commission Anti-Fraud Knowledge Centre 2021).

In the external or outsourcing model, analytical capability is developed by universities, thinktanks, private firms or civic-tech CSOs, while ACAs retain investigative and sanctioning powers. In practice, most ACAs operate across all three models at once, using in-house tools where they have direct access to structured data,

---

<sup>7</sup> These categories were derived through a synthesis of the comparative literature and documented country case studies on how anti-corruption authorities organise data access and analytical capability, including policy reports, and practitioner/ACA publications reviewed for and cited throughout this study.

collaborating with other state bodies for sensitive or specialist datasets, and selectively drawing on external analytical capacity.

## Cross-institutional data sharing

Making effective use of any of the big datasets for anti-corruption work depends heavily on cross-institutional data sharing. For ACAs, the most relevant datasets – procurement, tax, customs, corporate and beneficial ownership registers, asset declarations – are almost always held by other institutions (World Bank 2020). As a result, the quality of an ACA’s analytics is largely determined by the legal, technical and organisational arrangements that govern how information moves across government institutions (Transparency International 2023b).

### Drivers of data sharing

Several forces contribute towards greater data sharing between ACAs and their partners. First, functional necessity: complex corruption schemes typically cut across institutional boundaries, involving procurement authorities, line ministries, state-owned enterprises, tax administrations, regulators, FIUs and courts. No single body holds all relevant information, so collaboration is indispensable if patterns of bid-rigging, asset concealment or conflict of interest are to be detected at scale (World Bank 2020).

Data sharing also improves efficiency and coherence. Using existing registries and platforms is cheaper and more consistent than building parallel systems. For example, it is generally more efficient for an ACA to obtain a secure interface to the national e-procurement platform than to require separate reporting of contract data to its own database. The World Bank’s global review of digital anti-corruption tools further stresses that anti-corruption analytics work best when institutional users can link it with other data sources, rather than proliferating bespoke systems (World Bank 2020).

Third, there are normative and legal expectations. The UNCAC explicitly calls for “effective and coordinated” preventive policies and information exchange, and the UNODC guidance under Chapter II emphasises that specialised bodies should have timely access to public-sector information needed to discharge their mandates (United Nations Office on Drugs and Crime 2004; United Nations Office on Drugs and Crime 2020a). In parallel, the UNODC’s *Statistical Framework to Measure Corruption* is an ongoing initiative that seeks to strengthen the coordination in how countries define, measure, use and share corruption-related data across institutions (United Nations Office on Drugs and Crime 2023:3).

In practice, provisions like these require states to embed data access rights and obligations in domestic law, enabling ACAs to identify, analyze and address vulnerabilities before violations occur. Recent OECD and EU-level analyses of asset and interest declaration systems highlight the importance of legal and technical arrangements that allow oversight bodies to access external data sources (such as tax, banking, company and land registers), subject to appropriate safeguards, in order to verify declarations and detect risks more effectively (OECD 2023). For instance, Romania's ANI has legal powers to access fiscal registries, asset declarations, land registry, real estate registry as well as other property registers (GRECO 2023: 35).

## Barriers to data sharing

Despite these enabling factors, there are legal and logistical barriers for ACAs to access relevant big data through cross-institutional sharing arrangements.

The first challenge is legal fragmentation and privacy constraints. Sector-specific secrecy rules (for tax or banking information) combined with modern data-protection regimes, can restrict information sharing or make it procedurally burdensome (World Bank 2020: 230). In several countries, laws allow broad data access for criminal investigations but are more restrictive for preventive risk analysis. For example, access to banking data in Croatia is only permissible in the case of an ongoing criminal investigation (Hoppe 2013: 12). By comparison, Albania and North Macedonia waive bank secrecy for certain individuals, such as public officials, to facilitate the verification of asset declarations (Hoppe 2013: 12). Recent litigation on beneficial ownership registers in the EU, culminating in the Court of Justice of the European Union (CJEU) 2022 judgement limiting full public access to BO data, illustrates how privacy concerns can significantly reshape access regimes and require more targeted, "legitimate interest" based models (CJEU 2022; Open Ownership 2022). This means that stakeholders need to demonstrate and justify a legitimate objective when requesting the data (CJEU 2022). In practice, the various European countries interpret this differently, resulting in fragmented access to BO data across settings, further limiting the possibility of cross-country comparisons (Transparency International 2023c).

Technical incompatibilities pose a further barrier to effective data sharing. Data-holding bodies often use different identifiers for persons and entities, store information in legacy formats and lack secure, documented interfaces for machine-to-machine exchange. Recent studies on corruption-risk indicators in procurement and on risk-based asset-declaration systems underline that interoperable identifiers and minimal common data standards are preconditions for scalable analytics; without them, linking datasets is time-consuming and error-prone (Fazekas and Kocsis 2020).

Another issue to consider is the “tools - capacity gap”. Legal and technical reforms have delivered new datasets and sophisticated tools faster than ACAs can build the organisational and human capacity to catch up effectively. ACAs may be provided with access to e-procurement systems, BO registers and e-declaration platforms, but lack the staff, skills and workflows to turn these into actionable intelligence (OECD 2022b). A prominent example is Armenia’s *Կոռուպցիայի կանխարգելման հանձնաժողով* (Corruption Prevention Commission, CPC), mandated to verify extensive asset and interest declarations. OECD and World Bank assessments note that, until very recently, the CPC had limited staff and IT capacity; it relied on simple spreadsheets and manual checks, even as a new asset-declaration platform with interoperability features was being developed (OECD 2022b 5:6; World Bank 2023b). These inefficiencies might limit the ACA from engaging in and benefiting from effective cross-institutional data sharing.

## Emerging practices

To address some of these barriers, states can experiment with various mechanisms to enable cross-institutional data sharing while managing risks. Collaboration is also sometimes facilitated through joint analytical units and taskforces. As discussed, joint intelligence centres, sometimes hosted in one institution, bring together analysts from ACAs, FIUs, tax administrations, police and audit institutions to work with shared tools and datasets (Gunn and Scott 2018: 88). A World Bank review of inter-agency collaboration documents such joint risk-analysis units in revenue and customs administrations that systematically draw on data from anti-corruption bodies and vice versa (World Bank 2021).

In some countries, data exchange happens through shared technical infrastructures: secure data hubs or interoperability platforms that allow different authorities to access, link and analyse datasets under agreed rules, often using common identifiers. Experience from procurement risk platforms and AID systems in Europe shows that such hubs can simultaneously improve data quality, standardisation and analytical capacity across government (World Bank 2020; Di Nicola et al. 2025).

Even where these arrangements exist, they require continuous maintenance and political support. Leadership changes, budget cuts or high-profile scandals can quickly erode trust and stall cooperation, underscoring that cross-institutional data sharing is as much a governance project as a technical one.

## Risks associated with using big data

The emerging literature also highlights the potential risks of using big data and the tools needed to analyse it: problems of data quality and integrity (European Parliament 2021: 11-12) and the lack of transparency in AI mechanisms (Kossow et al. 2021: 14). These risks are not only technical in nature; they affect due process and institutional legitimacy.

The first category of risks concerns data quality. As laid out in previous sections, big data tools rely on the aggregation of heterogeneous administrative data, such as procurement awards, BO registers, asset and interest disclosures, etc. The accuracy of any analytical results is constrained by the quality and coverage of these inputs. In practice, for instance, information is often missing from procurement award notices. However, it is difficult to determine whether this is the result of low integrity in the procurement process or of a lack of information storage capacities and poor maintenance (Poltoratskaia and Fazekas 2024: 52). Moreover, databases can be difficult to link as matching the same entity across different systems is often hampered by missing identifiers or non-standardised formats (European Parliament 2021: 15). As a result, the poor quality of underlying data can translate into false positives (flagging low-risk actors) and false negatives (missing well-connected high-risk entities), undermining both the efficiency and credibility of ACA work. To avoid systematic bias, it is important to also understand what data is used to train analytical models and how. When researchers generate data based on biased assumptions or flawed measurement practices, and that data is used to train advanced analytical tools such as AI models, those models will potentially reproduce and amplify the same biases in their outputs (Kossow et al. 2021:5).

A second group of risks relates to the opacity of AI mechanisms and the implications for accountability. Many of the more sophisticated tools used or proposed for anti-corruption work function as partially opaque systems. The literature on algorithmic transparency stresses that computer algorithms in public decision-making can be biased both by the input data and by developers' design choices, and that machine-learning systems, in particular, often operate as "black boxes" whose inner workings are difficult to explain (Kossow et al. 2021). For instance, the Brazilian system ALICE demonstrated measurable results in tracing red flags in public procurement. However, the criteria and weightings behind ALICE's alerts are not fully transparent to suppliers or even internal users, raising accountability concerns (Neves et al 2019). Reviews of governmental practice find that such opacity exists across sectors, often intentionally to conform to legal standards (Valderrama, Hermosilla, & Garrido 2023: 3). In the anti-corruption field, Zinnbauer (2025: 10) argues rather than as a basis for automated decision-making, AI should be conceived primarily as a tool for supporting decisions made by humans who can be held accountable.

# Conclusion

Across different country contexts, data-driven tools can significantly strengthen ACAs' efforts in corruption prevention and detection. However, the literature emphasises that these novel approaches have to be built on solid legal frameworks, interoperable data and sustained institutional capacity. BO registers, income and asset declarations, procurement data, tax and company records, and related open data can reveal conflicts of interest, red-flag high-risk firms and support investigations.

The comparative experiences reviewed here – from Brazil to Indonesia, Ukraine and beyond – show that reforms work best when they are grounded in a comprehensive step towards digitalisation. The lack of fully developed data infrastructure, skills, and common data standards can prevent ACAs from reaching their full analytical potential. In practice, this shows up as uneven data quality, weak interoperability across government systems, and limited transparency and accountability around the data and analytical tools being used.

Ultimately, big-data tools should not be seen as a technical add-on but as a catalyst for transforming how ACAs understand and address corruption risks. When grounded in strong legal frameworks, interoperable systems and sustained institutional capacity, these tools can help ACAs move beyond isolated cases to counter systemic vulnerabilities, support evidence-based reforms and make anti-corruption efforts more preventive, targeted and resilient over time.

# References

Adam, I. and Fazekas, M. 2021. Are emerging technologies helping win the fight against corruption? A review of the state of evidence. *Information Economics and Policy*, 57, p.100950.

Arista, I. T., Fazekas, M. & Volkotrub, A. 2024. Using Beneficial Ownership Data for Large-Scale Risk Assessment in Public Procurement: the Example of 6 European Countries.

Government Transparency Institute Working Paper.

Autorità Nazionale Anticorruzione. No date. Database: Analytics Database and ANAC Open Data Portal.

Autorità Nazionale Anticorruzione. 2021. Measuring Corruption Risk At Territorial and Sectoral Level. Rome, Italy: ANAC.

Autorità Nazionale Anticorruzione. 2023. “Nuovo Codice Appalti, Pnrr, Banca dati, digitalizzazione: corso dell’Ordine dei Giornalisti con Anac”.

Bosisio, A., Carbone, C., Jofre, M., Riccardi, M. & Guastamacchia, S. 2021. Developing a Tool to Assess Corruption Risk Factors in Firms’ Ownership Structures: Final Report of the DATACROS Project. Transcrime – Università Cattolica del Sacro Cuore.

Burdescu, R., Reid, G. J., Gilman, S. & Trapnell, S. 2009. Income and Asset Declarations: Tools and Trade-Offs. Stolen Asset Recovery (STAR) Initiative) The World Bank Washington, DC.

CJEU. 2022. Joined Cases C-37/20 and C-601/20, WM and Sovim SA v Luxembourg Business Registers.

DATAcros I Consortium. No date.

DATAcros I – Partners and Actions (project website).

DATAcros III Consortium. No date.

DATAcros III – Partners.

Department for Business and Trade. 2025. People with Significant Control (PSCs). GOV.UK.

Department for Business, Energy & Industrial Strategy. 2019. Post-implementation review of the People with Significant Control Register (SI 2017/694).

Di Nicola, A., Baratto, G., Donati, B. & Rigon, B. 2025. Asset and Interest Disclosure (AID) Systems in EU Member and Candidate States: Report on the Standardized EU Risk Analysis Framework and Roadmap for Implementing Automated and Digital AID Systems (qAID Deliverable 3.1). Centre for Security and Crime Sciences (CSSC), University of Trento and University of Verona.

European Commission Anti-Fraud Knowledge Centre. 2020. PREVENT System. Good Practice Case Study.

European Commission Anti-Fraud Knowledge Centre. 2021. Special Investigation Service (STT), Lithuania.

European Parliament & Council of the European Union. 2015. Directive (EU) 2015/849 of the European Parliament and of the Council of 20 May 2015 on the Prevention of the Use of the Financial System for the Purposes of Money Laundering or Terrorist Financing, Amending Regulation (EU) No 648/2012 of the European Parliament and of the Council, and Repealing Directive 2005/60/EC of the European Parliament and of the Council and Commission Directive

2006/70/EC. Official Journal of the European Union, L 141, 73–117.

Fazekas, M., Tóth, I.J. and King, L.P. 2016. An Objective Corruption Risk Index Using Public Procurement Data. European Journal on Criminal Policy and Research, 22(3), pp.369-397.

Fazekas, M. and Kocsis, G. 2020. Uncovering High-Level Corruption: Cross-National Objective Corruption Risk Indicators Using Public Procurement Data. British Journal of Political Science, 50(1), 155-164.

Felbermayr, G., Kirilakha, A., Syropoulos, C., Yalcin, E. & Yotov, Y. V. 2020. The Global Sanctions Data Base. European Economic Review, 129, 103561.

Financial Action Task Force. (n.d.). FATF glossary. FATF-GAFI. <https://www.fatf-gafi.org/en/pages/fatf-glossary.html>

GRECO. 2023. Fifth Evaluation Round – Evaluation Report on Romania, Council of Europe.

Gunn, B. and Scott, E. 2018. Improving Co-operation between Tax Authorities and Anti-Corruption Authorities in Combating Tax Crime and Corruption. Washington, DC: World Bank Group.

Hoppe, T. 2013. Income and Asset Declarations in Practice: Comparative Study. Regional School of Public Administration (ReSPA).

Imazon. 2023. System for Monitoring Timber Harvesting (Simex): Mapping Logging in Brazilian Amazon – August 2021 to July 2022. Imazon, Idesam, Imaflora, & ICV.

Johnsøn, J., Heckler, H., De Sousa, L. and Mathisen, H. 2011. How to Monitor and Evaluate Anti-Corruption Agencies: Guidelines

for Agencies, Donors and Evaluators

U4 Issue 2011:8.

Kossow, N., Windwehr, S. & Jenkins, M. 2021. Algorithmic Transparency and Accountability. Anti-Corruption Helpdesk Answer. Transparency International.

Longobucco, A. & Ferwerda, J. 2025. Fighting Corruption with Artificial Intelligence: Searching for Suitable Public Procurement Data in the EU. Bridgegap Working Paper Series, 25(01). Utrecht: Utrecht University.

Maslen, C. 2025. Anti-Corruption Agencies and Public Engagement. Bergen: Transparency International and U4 Anti- Corruption Resource Centre, Chr. Michelsen Institute. U4 Helpdesk Answer 2025.

Moneyval. 2018. Anti-Money Laundering and Counter-Terrorist Financing Measures: Latvia – Fifth Round Mutual Evaluation Report. Council of Europe, Committee of Experts on the Evaluation of Anti-Money Laundering Measures and the Financing of Terrorism.

NACP (National Agency on Corruption Prevention). 2023. Electronic Declaration System in Ukraine: Verification and Risk-Rating Procedures.

Neves, F. R., da Silva, P. B. & de Carvalho, H. L. M. 2019. Artificial Ladies against Corruption: Searching for Legitimacy At the Brazilian Supreme Audit Institution. Revista de Contabilidade e Organizações, 13, 31–50.

OECD. 2022a. Modernising Integrity Risk Assessments in Brazil: Towards a Behavioural-Sensitive and Data-Driven Approach. OECD Public Governance Reviews. OECD Publishing.

OECD. 2022b. Anti-Corruption Reforms in Armenia. Paris: OECD.

OECD 2023. [Review of the Asset and Interest Declaration System in Malta](#). Paris: OECD.

OECD. 2025. [Developing a Data-Driven Corruption Risk Model to Strengthen Integrity in Belgium: Practical Considerations for the Federal Internal Audit](#). Paris: OECD Publishing.

Odilla, F. 2023. [Bots against Corruption: Exploring the Benefits and Limitations of AI-Based Anti-Corruption Technology](#). Crime, Law and Social Change, 80(4), 353–396.

Open Ownership. 2022. [Statement on the CJEU Judgment on Public Beneficial Ownership Registers in the EU](#).

Open Ownership. 2024. [Beneficial Ownership Transparency: A Fundamental Reform to Accelerate Sustainable Development](#).

OpenOwnership Briefing.

Parliament of Romania. (2016). *Law no. 184/2016 of 17 October 2016 to establish a mechanism to prevent conflict of interests in public procurement contract awarding*. Official Journal of Romania, Part I, no. 831, 20 October 2016. <https://integritate.eu/wp-content/uploads/2024/03/Law-184-2016.pdf>

Poltoratskaia, V. & Fazekas, M. 2024. [Data Analytics for Anti-Corruption in Public Procurement](#). In *Routledge Handbook of Public Procurement Corruption* (pp. 42-59). Routledge.

ProACT. No date. [Procurement Anticorruption and Transparency Platform](#).

Purcell, J. & Rossi, I. 2019. [Privacy Vs Transparency: Countries Must Strike the Right Balance As They Combat Illicit Financial Flows](#). Finance & Development, 56(3), 18–21.

Putra, Y. R. P. 2024. [Diving Deeper to the State Administrator Asset Report \(LHKPN\): From](#)

Public Perception to Data Analytics. Asian Public Finance Journal, 9(1).

Sagiroglu, S., and Sinanc, D. 2013. [Big Data: A Review](#). In *2013 international conference on collaboration technologies and systems (CTS)* (pp. 42-47). IEEE.

Schütte, S., Ceballos, J. C. and David-Barrett, E. 2023. [Measuring Effectiveness of Anti-Corruption Agencies: Robust, Relevant and Useful Measurement](#). Insights Briefs Series No. 07 / August 2023.

Schütte, S. A. and David-Barrett, E. 2025. [Assessing Compliance with the Jakarta Principles: A Grading Framework](#). Bergen: U4 Anti-Corruption Resource Centre, Chr. Michelsen Institute (U4 Brief 2025:4).

Transparency International Romania. 2022. [Boosting Integrity in Romanian Public Procurement](#).

Transparency International Ukraine. 2022. [NABU Serves with Charges in Case that Began with Feedback on DOZORRO](#).

Transparency International. 2023a. *UNCAC CoSP10: Access and use of beneficial ownership registers*. Transparency International. <https://files.transparencycdn.org/images/UNCAC-CoSP10-Access-and-Use-of-Beneficial-Ownership-Registers.pdf>

Transparency International. 2023b. [Anti-Corruption Agencies and Income, Interest and Asset Declaration \(IIAD\) Systems: An Overview of Good Practice Examples](#). Transparency International Anti-Corruption Helpdesk Answer

Transparency International. 2023c. [Access to Beneficial Ownership Registers Under EU AMLD6 and the CJEU Ruling: Policy Position](#)

(Policy position paper). Transparency International.

United Nations Department of Economic and Social Affairs. 2024. [CEPA Strategy Guidance Note on Registries of Beneficial Ownership](#). New York: United Nations Department of Economic and Social Affairs, Division for Public Institutions and Digital Government.

United Nations Office on Drugs and Crime (UNODC). 2004. [United Nations Convention against Corruption](#). United Nations.

United Nations Office on Drugs and Crime. 2012. [Jakarta Statement on Principles for Anti-Corruption Agencies](#).

United Nations Office on Drugs and Crime (UNODC). 2019. [Income and Asset Declaration Systems in Indonesia. Country Analysis](#).

United Nations Office on Drugs and Crime (UNODC). 2020a. [Colombo Commentary on the Jakarta Statement on Principles for Anti-Corruption Agencies](#). Vienna: UNODC.

United Nations Office on Drugs and Crime. 2020b. [Asset Tracing Country Profile: Indonesia](#). G20 Anti-Corruption Working Group.

United Nations Office on Drugs and Crime. (2023). *Statistical framework to measure corruption*. [https://www.unodc.org/documents/data-and-analysis/statistics/corruption/UNODC\\_Statistical\\_Framework\\_to\\_measure\\_corruption.pdf](https://www.unodc.org/documents/data-and-analysis/statistics/corruption/UNODC_Statistical_Framework_to_measure_corruption.pdf)

U4 Anti-Corruption Resource Centre. 2023. [The Use and Impact of Beneficial Ownership Information in Anti-Corruption](#). U4 Helpdesk Answer.

Valderrama, M., Hermosilla, M. P. & Garrido, R. 2023. [State of the Evidence: Algorithmic](#)

[Transparency](#). GobLab UAI & Open Government Partnership.

World Bank. 2017. [Guide to Beneficial Ownership Information in Ukraine: Legal Entities and Legal Arrangements](#). World Bank.

World Bank. 2020. [Enhancing Government Effectiveness and Transparency: The Fight Against Corruption](#). World Bank, Washington, DC.

World Bank. 2021a. [Reform of Asset and Interest Disclosure in Ukraine](#). Case Study 18 in Enhancing Government Effectiveness and Transparency: The Fight Against Corruption.

World Bank. 2021b. [Inter-Agency Collaboration to Detect Corruption](#). Washington, DC: World Bank.

World Bank. 2023a. [Asset and Interest Disclosure: A Technical Guide to an Effective Form](#). World Bank Group & UNODC StAR Initiative.

World Bank. 2023b) [Strengthening Public Sector in Armenia - Inputs for a Digitalization Strategy for Corruption Prevention](#).

Zinnbauer, D. 2025. [Artificial Intelligence in Anti-Corruption: A Timely Update on AI Technology](#). U4 Brief 2025:1. U4 Anti-Corruption Resource Centre, Chr. Michelsen Institute.

---

## Disclaimer

All views in this text are the author(s)', and may differ from the U4 partner agencies' policies.

---

## Creative commons

This work is licenced under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0

International licence (CC BY-NC-ND 4.0)



**Corruption erodes sustainable and inclusive development. It is both a political and technical challenge. The U4 Anti-Corruption Resource Centre (U4) works to understand and counter corruption worldwide.**

**U4 is part of the Chr. Michelsen Institute (CMI), an independent development research institute in Norway.**

[www.u4.no](http://www.u4.no)

[u4@cmi.no](mailto:u4@cmi.no)

---

## U4 partner agencies

German Corporation for International Cooperation – GIZ

German Federal Ministry for Economic Cooperation and Development – BMZ

Global Affairs Canada

Ministry for Foreign Affairs of Finland

Ministry of Foreign Affairs of Denmark / Danish International Development Assistance – Danida

Norwegian Agency for Development Cooperation – Norad

Swedish International Development Cooperation Agency – Sida

Swiss Agency for Development and Cooperation – SDC

UK Aid – Foreign, Commonwealth & Development Office