

# INTEROPERABILITY: LINKING UP DATA AND PEOPLE

## Introduction

The evidence is growing: the right data, paired with robust oversight, can make a difference. [Combining](#) each [dataset](#) with others, however, increases the [effectiveness](#) of any single dataset (Rossi, 2012; Arce, Falla, & Foti, 2019; Global Financial Integrity, 2022). For example, lobbying data is important. When combined with political finance data and beneficial ownership data, however, lobbying data becomes even more powerful for tracing the flow of not only ideas, but also money in politics. Similarly, open contracting data has been shown to have numerous benefits; when joined with political finance data, it can help detect potential kickbacks or opportunities for illicit enrichment. This section looks at what steps governments can take to ensure that data to counter political corruption can be joined up to ensure this synergy.

This analysis is part of the [Broken Links: Open Data to Advance Accountability and Combat Corruption](#) report, which offers an overview of data frameworks and data availability in OGP countries across eight policy topics using data from the Global Data Barometer.

## Key Takeaways

**Human and machine interoperability:** Getting people together and starting small is essential to success, whether or not the data collection and publication systems are up to standard. Getting people to talk, especially across sectors, can be key to making sure that anti-corruption data makes politics cleaner and more fair.

**Reform is necessary and overdue:** Using standardized, validated approaches to data collection and publication makes individual datasets exponentially more powerful in combating corruption. Lobbying data is more powerful when asset declarations are made public, and procurement data works better when the public knows who owns winning companies.

**Common identifiers:** According to the Global Data Barometer (GDB), most OGP countries do not have common identifiers when it comes to companies, legislation and regulation, politicians, lobbyists, or other essential data. Given how much work remains for most OGP countries, building validated common identifiers into the system early will be essential.

Open County Kilifi monitors and evaluates the performance of county projects. Open Institute worked with the county government and young citizens to collect data on the status of these projects. This initiative also includes a review phase where supervisors can verify the date in real time and flag where clarification is needed or data quality is low. Overall this project is serving to enhance citizen engagement and ownership by its own citizens. Photo by Open Institute.



Open  
Government  
Partnership

# Policy Justification

There are strong arguments for interoperability beyond the control of corruption. Democratic political systems are complex, with many actors, many organized interests, and consistent, significant government activity. The following are arguments for making data interoperable (see *Good to Know: Key Terms*).

- **Corruption often doesn't involve only a single act, type of act, or actor, but rather entails networks and flows.** Data can be a critical tool in tracking illicit financial flows and otherwise fighting corruption, but when the relevant data types aren't interoperable, it may offer only a fragmentary picture. However, making such data interoperable—for example, using the same unique identifiers across different types of datasets—makes it increasingly useful.
- **Interoperability allows people to learn and compare across jurisdictions, globally, and within a given country.** This is essential for learning but also comparison and, in some cases, may be useful for legal and trade harmonization. An example of where national comparability of asset disclosure, for example, would be valuable is ensuring that bribe paying is not occurring to ministries or legislators by multinational corporations or locally based companies. This levels the playing field to some extent. Similarly, investors may want to understand the political risk their companies or bonds are subject to when they invest in countries with weaker accountability institutions.

- **Interoperability allows people to trace influences between levels of government, especially where a government function is federalized or decentralized.** Being able to compare provincial, regional, or state-level data allows national governments to identify innovators and areas for capacity improvement at the state level. This can be especially important where spending capacity on infrastructure, health, or education is delegated to lower levels of government.
- **Perhaps most importantly, interoperability is essential for data to be useful for different sectors of society—including government regulators, the press, watchdog groups, the private sector, and interested citizens.** Ensuring that data is standardized allows these organizations to recombine it with other datasets and information at a speed and rate that is otherwise impossible without some standardization for interoperability.

## GOOD TO KNOW

### Key Terms

Surprisingly, there are few resources defining interoperability in the space of democracy and control of corruption. This report borrows the definition from the United States Federation of Government Information Processing Councils, which defines it as:

*Interoperability of systems only requires a common basis for those elements that are, in fact, shared. Typically, not all of the information managed by two systems is shared. Therefore, interoperability requires identifying the **shared elements**. Not even all elements that have common (or close) definitions need to be shared. **Interoperability involves common semantics and syntax only for those elements that must be combined, compared or aggregated.** [Emphasis added.]*

For purposes of this report, those elements or common identifiers which need to be combined, compared, and aggregated are laid out in detail in Table 1: *Common identifier by justification*. They include: officials; interests, assets, and liabilities; legal persons (including companies); and identifiers for legislation and regulation.

**Common identifiers** are specific data fields or sets of data fields that allow for more automated approaches to information exchange. By standardizing these common identifiers, they allow for: better information flow across systems (e.g., between a database with lobbying clients and company ownership); decentralized management of data in different parts of government; and a reduction in the need for duplicate copies of information about important subjects (such as people, politicians, or countries).

Experts interviewed in the production of this report also suggested that, in many contexts, data may simply not yet be standardized enough. Given this potential limitation, they encouraged the report to also cover “**human interoperability**,” wherein individuals in key agencies (or outside of government) meet to ensure that information can still be exchanged, even if this information is in documents, rather than available as structured data. One expert cited the example of the relative success of the Nigerian Extractive Industry Transparency Initiative as a model worth exploring. While the multistakeholder forum responsible for reporting has been unable to produce uniform data in a central database, it has been able to regularly bring together 12 distinct agencies to harmonize key data on the country's oil and gas sector for a number of years. This has been a noteworthy improvement (Hatcher-Mbu, 2022); earlier data had largely been unavailable or inconsistent across sources.





In the Republic of Moldova, the public budget is largely misused through fraudulent and corrupt public procurement procedures. While the country does have an e-procurement system called MTender where information on the procurement process is available, only experts in public procurement can properly interpret the data and draw truthful conclusions. Recently, reformers from the Association for Efficient and Responsible Governance (AGER) launched the monitoring portal Revizia.md to help ordinary citizens understand the most frequent irregularities. They hope to encourage civil society groups and media to detect and report fraud and demand better accountability from their government. Photo by AGER.

# Analog and Digital Infrastructure

Linking up political data requires more than collecting and publishing data. It also requires “analog” or “human” interoperability. This means making sure that people and organizations have the relationships and ability to unlock the data. It also means having institutional structures that make collaboration easier and more effective than working in silos. Eventually, this means of interoperability similarly requires development of common identifiers and data standards to make everything work smoothly and reliably.

Given that linking up data is as much about humans as it is about data standards, reformers seeking to capitalize on the data may choose different sequences. Some may take a “standards-first” approach, while still others focus on bringing people and organizations together and letting common

practices evolve, a “people first” approach. Which strategy to take probably depends on the strength of the legal mandate for linking up data, the age of existing systems, and the complexity of the institutions responsible for collection and publication of the data. In cases where capacity, data quality, or cooperation are low from the outset, it may make sense to opt for a “people first” approach. *Lessons from Reformers: Getting “Human Interoperability” Right in Kenya* takes a look at how, in the absence of more technical approaches, journalists and watchdog groups are still bringing together data to unearth corruption.

Acknowledging that some governments may wish to take the “analog” approach, it is still worth looking at what is known about more standardized approaches. Here the GDB data gives us an insight into one measurable aspect of implementation: the existence of common identifiers.

## LESSONS FROM REFORMERS

### Getting “Human Interoperability” Right in Kenya

While many of the findings of the GDB focus on the issues of common identifiers and the systems that would support linking up data, it is worth taking some time to look at cases where standardization may not be immediately attainable. Lessons from big reforms in Kenya show some of the ways in which the reform process can move forward even in the absence of a high-tech approach.

#### The Case of Kenyan Contracts

Open contracting has been an area of major focus in Kenya. Progress on contracts is difficult, however, precisely due to issues of interoperability. In this particular case, there is a lack of standardization between the national and county-level procurement data publishing portals. As a start, data in both national and county-level procurement portals is mapped to the Open Contracting Data Standard (OCDS), creating an opportunity for standardization. Secondly, in May 2022, the open contracting cluster (consisting of both county and national governments), under the Open Government Partnership, developed a road map to address both policy and technical aspects of interoperability in the coming two years.

A number of procurement processes take place at the county level. The [Public Procurement Regulatory Authority](#) (PPRA, n.d.) has [statutory authority](#) to create regulations and guidance, as well as to collect relevant data (PPRA, 2019). The PPRA does not, however, have powers of enforcement for when procuring authorities do not collect or publish their data. Further, it lacks a direct mandate to standardize and collect all of the data from the local level. Consequently, many counties either do not submit or store data in their own, proprietary software. Because the PPRA cannot independently collect, standardize, or collect data, it is left with an enforcement problem.

Without adequate enforcement, standard procedures, data, and transparency, opportunities for corruption and “informal arrangements” multiply. This is best documented by the investigative journalism organization Africa Uncensored. In a series of investigations entitled

*Captured* (2019–2021), journalists, using leaked documents from the Integrated Financial Management System, found serious conflicts of interest in the awarding of contracts to private companies with no expertise (Africa Uncensored, 2019).

Currently, because there is inadequate data and little consequence for nonreporting, journalists and activists must rely on unofficial data to report on issues of public interest. Government units working to regularize procurement, as well, often rely on informal means of enforcement (such as leaks) rather than following straightforward legal channels. In an ideal scenario, the data would be available, complete (covering all procuring authorities), comparable, and would allow people to investigate the tenders, the awards, the companies (and their owners), and the status of implementation.

In the coming years, Kenya will roll out a new financial system, the e-GP (or e-Government Procurement). All counties and national government entities will be mandated to procure through this system. Ideally, such a system should also be able to support disclosure to the entire public.

#### **Hidden Owners**

Since the London Anti-Corruption Summit in 2016, President Uhuru Kenyatta has publicly pursued the goal of publishing beneficial ownership data. At the same time, there have been numerous scandals that have plagued the administration, most notably around health contracts (during COVID-19) and security contracts. One specific issue is the awarding of contracts to unqualified shell companies, companies that have been banned but have re-registered, or cases where the same company applies multiple times through different shell companies. These issues could be better dealt with if there were clear, publicly available beneficial owners.

Beneficial ownership data, however, remains less useful than it might otherwise be. To improve the utility of the ownership data, it needs to better align with data on politically exposed persons or asset disclosure data. Further, conflict of interest laws that apply to “public officers” or civil servants apply in a more limited number of cases to “state officers,” a different category that includes parliamentarians.

In the absence of this data, or where data is messy, there is often a reliance on nongovernment actors—whether media, NGOs, or private companies.

There is, however, currently little formal or regular coordination between the government and journalists or other watchdog organizations. While there is a multistakeholder working group on corruption that meets regularly, it has been limited in impact. Additionally, its minutes are not public. Nonetheless, the group can informally refer cases or findings to other processes for investigation or prosecution.

#### **Avenues of Accountability**

There are several other avenues to human interoperability that experts say might help Kenya.

The first would be to bring in private sector and finance sector allies that wish to help create a more level playing field to form official partnerships with government champions of corporate transparency. Indeed, the International Monetary Fund has been working to make ownership data for contractors public. A number of multinational companies have also come out in favor of open ownership. Bringing these powerful players into the discussion will be necessary, as there are strong forces working against open ownership information—from local companies resisting regulation to banks that are unaccustomed to customer due diligence and lawyers arguing for client anonymity.

The second avenue in the absence of strong regulatory capacity would be to strengthen social accountability, which has been done recently by the Office of the Auditor General. The Auditor General has asked for public information from organized groups of citizens through a structured social audit process in comparison with infrastructure.

The third avenue would be to strengthen public sector and private sector whistleblower protections and rewards. In other countries, these have been shown to be effective at incentivizing civil servants and private sector actors to identify cases of waste, fraud, and abuse, especially where there is a clear pathway to release information through official channels.

# Common Identifiers

Assuming that, at some point, regulators, watchdog organizations, and parliaments want to make anti-corruption monitoring better, they will eventually want to introduce common identifiers. Common identifiers are a core part of the structure of a dataset and are, ideally, used across multiple datasets.

For the purposes of common identifiers for linking up anti-corruption, there is no single current data standard. Of course, existing data standards such as the Open Contracting Data Standard (OCDS) and Open Ownership can be easily joined in jurisdictions where both standards are adopted. In the absence of standardization, however, there are a number of identifiers that remain essential. The common indicators and their justifications are shown in Table 1.

TABLE 1: Common identifier by justification

COMMON IDENTIFIER TOPIC	DATASETS USING IDENTIFIER	JUSTIFICATION
Public Officials		
Public officials identifying information	Asset disclosure, lobbying, political finance	Most anti-corruption datasets require information on politicians. These include asset disclosure, lobbying, political finance, as well as election and voting records (not included in this dataset).
Interests, assets, and liabilities	Asset disclosure and political finance	Political parties, politicians, and donors should be able to identify accounts, expense types, and donations. This allows auditors and watchdog organizations to identify inconsistencies and identify instances of illegal gifts or spending.
Interests		
Lobbyist clients and party and campaign donors	Lobbying data and political finance	In addition to having unique identifiers for lobbying, unique identifiers for interest organizations (regardless of legal form) allow the reconciliation of lobbying (communication) with political finance (money).
Company ownership information	Lobbying data	Company ownership data is essential for identifying who is active in lobbying and owning land, as well as understanding when there are conflicts of interest with politicians.
Legal persons	Lobbying, political finance, and companies' data	Creating common identifiers for legal persons, including businesses, NGOs, and other such organizations ensures better understanding of who is seeking to inform government-decision-making as well as who is affected by it.

COMMON IDENTIFIER TOPIC	DATASETS USING IDENTIFIER	JUSTIFICATION
Corruption Vectors and Targets		
Land ownership	Political corruption data (general)	Open land ownership data is essential for identifying which interests lie behind political decisions, who benefits from public finance decisions, as well as major changes in land use. It is also essential to help track and verify whether real estate assets declared by candidates and appointees match public records.
Regulations and legislation	Lobbying registers and public consultation	Unique identifiers are essential to allow people to see which laws and regulations are the objectives of lobbying and public participation, whether in the legislative or executive branches of government.
Public finance	Political corruption data (general)	Common identifiers for budget items allow people to see where potential conflicts of interest may lie and who benefits from spending priorities.
Public procurement	Political corruption data (general) and public procurement	Common identifiers for contracts and bids allow oversight of kickbacks, ghost contracts, or potential conflicts of interest.

# Key Findings from the Global Data Barometer

GOOD TO KNOW:

## About GDB Data

The Global Data Barometer (GDB) is a global expert survey drawing on primary and secondary data that assesses data availability, governance, capability, and use around the world to help shape data infrastructures that limit risks and harms. Together with regional hubs and thematic partners, GDB researchers collected data on 109 countries, including 67 of the 77 OGP participating countries. The GDB captures data developments between May 1, 2019, and May 1, 2021, and includes 39 primary questions and over 500 sub-questions.

GDB researchers looked at the existence of common identifiers between different datasets. Researchers began by looking at data already located for political finance, interest and asset declarations, lobbying, and public consultation to see if they used shared fields of data between them. They also looked to see if common identifiers were present between the various datasets in this report such as companies ownership data, land ownership data, public procurement data, and public finance data. Having interoperability between these datasets speeds up how quickly journalists, oversight bodies, and independent watchdog organizations can find patterns across datasets. Figure 1 below visualizes how consistently different countries employ common identifiers across sets.

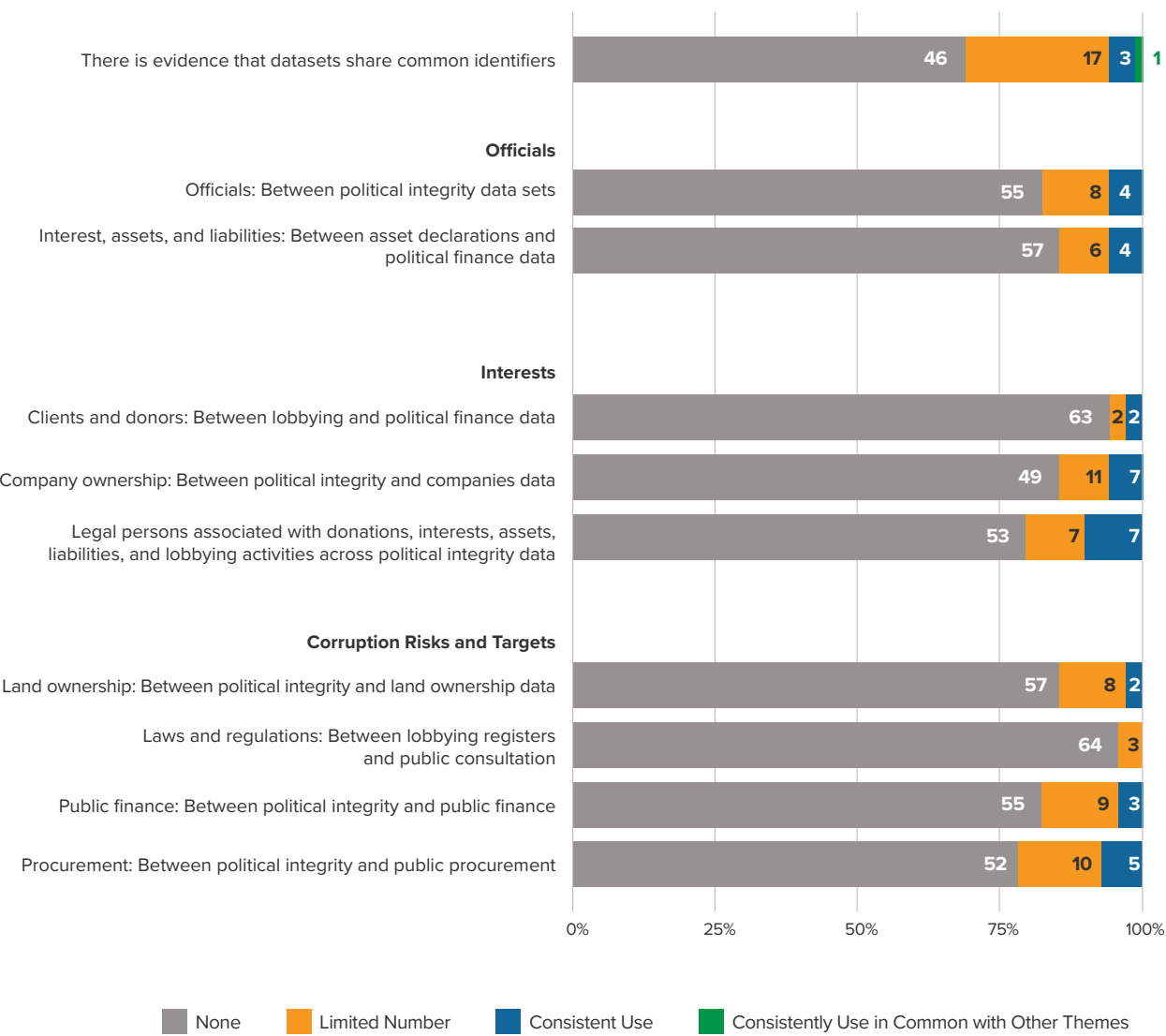
Highlights from the GDB data research and assessment include:

- **Absence of data is the main problem:** Most countries surveyed do not have common identifiers for the simple reason that they do not publish all of the relevant datasets. This cannot be stressed enough. It does, however, mean that as countries move to develop data entry, validation, and publication systems, they can do so knowing that it will save time and make data more useful to collect certain data in separate, reusable databases, rather than multiple times, in multiple ways.

- **Where data is available, there is some interoperability:** Where data is available, there is evidence that such sets do employ useful common indicators in at least some of the datasets (see Figure 1).
- **Companies’ data is harmonized between datasets most frequently:** Perhaps because it is one of the newest areas, company identifying data is frequently used across multiple datasets and policy areas. (see Figure 1).
- **Lobbying disclosure has low amounts of interoperability:** Disclosing lobbying data, in general, is rarely required and rarely disclosed. It thus follows that lobbying data is rarely disclosed in a way that allows for linking up data.

FIGURE 1. Prevalence of common identifiers across datasets

This figure shows the percentage of OGP countries with essential common identifiers. The sample includes all 67 OGP countries assessed by the GDB.





## Recommendations

The following recommendations illustrate a number of ways for governments to improve interoperability of data to counter political corruption, including through their OGP action plans. In addition, *Lessons from Reformers: Building Systems for High-Quality, Interoperable Data* offers strategies from professionals beyond policy change.

### Building Systems for Interoperability

- Ensure that fields in datasets reuse common identifiers for key pieces of information.
- Build data-entry systems to validate upon entry.
- Whole-of-government approaches may be needed to ensure that unique identifiers can be used across different datasets. (For example, ensure that company tax identification numbers are the same used for employment, permits, and addresses.)
- Some governments may have significant amounts of identifiers under legacy systems. Where legacy identifiers that should be common overlap but are not coterminous, consider phasing out legacy identifiers or creating “keys” for equivalence and interoperability. (See *Lessons from Reformers: Dealing With Legacy Common Identifiers in the United States*,” for example, on how the United States government has phased out Dun & Bradstreet numbers for Unique Entity Identifiers, which are public and relate directly to ownership, tax, and employment records.)
- Where data collection and publication systems are centralized, national governments can work with other countries to develop, adopt, and implement common standards.
- National governments can spur interoperability by convening data users and producers. This is especially important in federal systems where their national-level governments lack the ability to compel subnational governments but where subnational governments may benefit from information sharing, comparison, and learning across jurisdictions.

- Similarly, even between ministries and agencies, there may be more need to use common identifiers and reporting standards. This may require more explicit mandates for coordinating agencies as well as investment in software and data systems that make interoperability easier than inventing new systems.

### Toward Standard Common Identifiers

Because corruption, political influence, and organized crime are often cross borders, moving toward common identifiers will allow transboundary comparison. This is essential in tracking political finance flows. But it may also be useful in other, unexpected fields such as tracking transboundary pollutant transfers or fisheries usages.

- International standard-setting bodies can work together to ensure that they are working toward common identifiers and standardized systems of data. This could apply to regular OGP partners and affiliates such as the Open Ownership, Open Contracting Partnership, ISO standards bodies, and the various financial reporting organizations.
- Individual governments may convene stakeholders inside and outside of government to understand how well common identifiers do and do not work to solve particular, applied problems (such as tracking contract abuse).

### Human Interoperability

In the absence of open data and common identifiers, governments, donors, and watchdogs can take steps to promote the linking up of core data. Notable examples include:

- Convene potential users to prioritize common indicators.
- Build data structures with input from end users in a way that ensures that fields are consistent and useful in applied settings.
- Ensure adequate protection for whistleblowers acting in the public interest, especially where there is overwhelming interest (as in the Kenya case study above).

#### LESSONS FROM REFORMERS:

### Dealing With Legacy Common Identifiers in the United States

In the United States, a single factory or facility may have dozens of different identifiers when filing compliance data with the federal government. The system has been built up over decades, one law at a time. As a result, activists, journalists, or regulators looking into the operations of a single facility will need to know a variety of identifying numbers or codes.

Depending on the specific question at hand, this may include:

- Employer tax identification numbers when dealing with tax, employment, and worker safety issues
- Unique permit identification numbers for safety and environmental compliance data
- Data Universal Numbering System (DUNS) numbers when dealing with federal contracts
- Geospatial data
- Ownership data through Securities and Exchange Commission or state corporate registries

As part of its 2013 OGP national action plan ([Milestone 32.2](#)), the United States (n.d.) committed to moving toward a more interoperable and public data system. The US government proposed a rule in November 2015 to replace proprietary identification of entities with a generic terminology. This rule is part of an effort to move away from the DUNS, a standard developed by the private company Dun & Bradstreet that keeps track of entities receiving government funds. Given that the DUNS is a proprietary standard, much of the data on contractors is not public or reusable. For this reason, both the Government Accountability Office and leading open data groups have advocated for a move away from the DUNS. The [IRM found](#) this commitment to be noteworthy, both in design and implementation (OGP, n.d.).

As of April 22, 2022, the United States government has [officially retired DUNS numbers](#) and moved all contracting entities to a Unique Entity Identifier created in SAM.gov (US GSA, 2022). This allows for a single, publicly owned, nonproprietary common identifier for companies. While there is still room for further integration with other data systems, this takes a significant step toward retiring siloed, non-interoperable identifiers.

#### LESSONS FROM REFORMERS

## Building Systems for High-Quality, Interoperable Data

Open government reformers can draw from decades of work linking up data to counter political corruption. Replicable lessons cover issues related to lobbying, beneficial ownership transparency, open contracting, or asset disclosure, among others—and offer a more informed and useful understanding of who shapes policy and spending. A number of partners were interviewed in preparing this report. Common themes from these discussions include:

- Common identifiers are a core part of the structure of a dataset and are, ideally, used across multiple datasets. But the end goal is to achieve structured data. This means creating content fields that resonate with or are useful to data users. This requires regular communication with end users.
- Validation is essential to timely, high-quality data. And a centralized or standardized approach can ensure that data is validated. Examples cited include:
  - Public validation: Data is published, and stakeholders (regulators, regulated entities, and members of the public) can identify outliers, mistakes, etc.
  - Automated validation: A business, employer, or other official number is submitted, and it is validated through a central database.
  - Spot validation: Forms are written in a way that it may be impossible to register without validation, for example, requiring entry of beneficial owners' identification, which can be verified with another database.
- Dates are essential in determining when activities took place.
  - One form of fraud involves moving companies to other owners while applying for a bid, allowing a company to bid multiple times or for blacklisted firms to apply for a contract. Therefore, dates are essential to detect transfer of ownership to conceal personal interests.

- Quality assurance is also important. However, some means have proven more effective than others.
  - Publishing before a full audit is essential. Assigning an official to identify outliers and misreporting is important. However, holding back on publication *before* outliers are identified and resolved is not ideal, as it can result in undue delays in the release of data.
  - Early release of data should be combined with a *public* flagging system to encourage identification of suspicious activity and allow regulated entities to determine mis-entered data.
  - Consider releasing data that may be viewed as incomplete or imperfect. In some cases, especially with lobbying and asset disclosure, flawed data or concerns about privacy (sometimes made in good faith) need to be acknowledged and dealt with on a systematic and case-by-case basis.



## List of Resources

Africa Uncensored, “Captured: Stella’s Web,” YouTube Video, 17:45, September 27, 2019, <https://www.youtube.com/watch?v=ruh5QUzdWjo&list=PL8SSOlB-3zzK8xILT2JD5GJj4Zzt1SWOrk&index=4>.

Public Procurement Regulatory Authority, “Executive Order 2 of 2018,” 2019, <https://ppra.go.ke/download/executive-order-2-of-2018/>.

Public Procurement Regulatory Authority, “Big Infrastructure Projects,” n.d., <https://ppra.go.ke>.

Open Government Partnership, IRM staff, *Independent Reporting Mechanism (IRM): United States Progress Report 2015–2016*, n.d., [https://www.opengovpartnership.org/wp-content/uploads/2018/01/United-States\\_Mid-Term\\_2015-2017.pdf](https://www.opengovpartnership.org/wp-content/uploads/2018/01/United-States_Mid-Term_2015-2017.pdf).

United States, “Increase Transparency in Spending (US0039),” n.d., <https://www.opengovpartnership.org/members/united-states/commitments/US0039/>.

US General Services Administration, “Unique Entity Identifier Update,” 2022, <https://www.gsa.gov/about-us/organization/federal-acquisition-service/office-of-systems-management/integrated-award-environment-iae/iae-systems-information-kit/unique-entity-identifier-update#:~:text=On%20April%204%2C%202022%2C%20the,website%20to%20obtain%20their%20identifier.>

Interoperability: Linking Up Data and People is part of the *Broken Links: Open Data to Advance Accountability and Combat Corruption* report that can be found here: <https://www.opengovpartnership.org/broken-links/>