Mathematica
Progress Together

Independent Evaluation of the Kosovo Threshold Program Transparent and Accountable Governance (TAG) Project

Evaluation Design Report

December 21, 2022

Anthony Louis D'Agostino, Josh Meuth Alldredge, Hailey Hannigan, Matt Sloan, and Ksenia Miliutinskaia

Submitted to:

Millennium Challenge Corporation
1099 14th St NW, Suite 700
Washington, DC 20005
Project Officer: Kobby Amponsah
Contract Number: 95332419A0004

Submitted by:

Mathematica
1100 1st Street, NE, 12th Floor
Washington, DC 20002-4221
Phone: (202) 484-9220
Fax: (202) 863-1763

This page has been left blank for double-sided copying.

Contents

Tables

This page has been left blank for double-sided copying.

Figures

This page has been left blank for double-sided copying.

Acronyms

AQ	Air quality
CAPT	Contribution analysis with process tracing
CSO	Civil society organization
CTM	Case tracking mechanism
DQR	Data quality review
DUA	Data use agreement
EDC	Environmental Data Collection
ERO	Kosovo Energy Regulatory Office
FGD	Focus group discussion
GDP	Gross domestic product
GoK	Government of Kosovo
GSI	Gender and social inclusion
ICT	Information and communications technology
ITT	Indicator tracking table
KEPA	Kosovo Environmental Protection Agency
KHMI	Kosovo Hydrometeorological Institute
KII	Key informant interview
KJC	Kosovo Judicial Council
KLSC	Kosovo Legal Services Company
KODC	Kosovo Open Data Challenge
KPC	Kosovo Prosecutorial Council
KPP	Kosovo Public Pulse
MCC	Millennium Challenge Corporation
M&E	Monitoring and evaluation
MFK	Millennium Foundation Kosovo
MOJ	Ministry of Justice

NGO	Non-governmental organization
NIPH	National Institute of Public Health
ODP	Online data portal
OECD	Organisation for Economic Co-operation and Development
PAJI	Public Access to Judicial Information
PEA	Political economy analysis
PPP	Purchasing power parity
RELP	Reliable Energy Landscape Project
RMSE	Root mean square error
TAG	Transparency, Accountability, and Governance
TOC	Theory of change
UNDP	United Nations Development Programme
USAID	United States Agency for International Development
VPN	Virtual private network
WMS	Web map service

I. Introduction

Since Kosovo asserted its independence in 2008, its gross domestic product (GDP) has grown at an average rate of 4.2 percent per year, outperforming most Organization for Economic Co-operation and Development (OECD) countries. However, Kosovo's current GDP per capita of less than US$13,000 (in purchasing power parity, PPP) lags behind neighboring Balkan states like Albania, Serbia, Montenegro, and North Macedonia (World Bank 2022). In a recent analysis of key macroeconomic factors that limit Kosovo's economic growth, Zogaj et al. (2017) identified unreliable electricity supply, weak rule of law, and environmental pollution and poor air quality as among the most significant constraints to Kosovo attracting foreign investment and increasing citizens' trust in its democracy.

To address these challenges hindering Kosovo's economic development, the Millennium Challenge Corporation (MCC) partnered with the Government of Kosovo (GoK) on the Kosovo Threshold Program, a collaboration with a $49 million budget that entered into force in September 2017 and is scheduled to close in September 2022. The Threshold Program includes two projects, the Reliable Energy Landscape Project (RELP), designed to reduce the imbalance between energy demand and supply, and the Transparent and Accountable Governance Project (TAG), which aims to increase the availability and accessibility of data to the general public, encourage its use, and ultimately accelerate data-driven decision making.

The TAG project comprises three activities (Figure I.1): (1) the Public Access to Judicial Information (PAJI) activity, which aims to support the GoK's efforts to improve decision making and accountability by increasing the accessibility and public use of judicial data; (2) the Environmental Data Collection (EDC) activity, which aims to improve the quality and availability of environmental data to support data-driven decision making; and (3) the Kosovo Open Data Challenge (KODC) activity, which aims to foster a productive partnership between the GoK, the media, and civil society by supporting innovation in data use.

In September 2021, MCC contracted Mathematica to conduct an independent evaluation of the TAG project to determine how project activities and sub-activities contributed to improving the availability and use of public data and promoting a culture of transparency and data-informed policymaking. In the following chapters, we provide context for the evaluation and describe the planned evaluation design. In Chapter II, we provide a detailed description of the TAG project and its three activities, along with a discussion of the theory of change, which summarizes how project components contribute to achieving the intended outcomes. Chapter II also contains a literature review in which we discuss previous examples of open data initiatives governing judicial and environmental data to contextualize the key research contributions of this evaluation. We present our evaluation design in Chapter III, describing the research questions, methods, data sources, and key assumptions for evaluating PAJI, EDC, KODC, and the project as a whole. We also summarize our assessment of the evaluability of TAG and each activity separately in Chapter III, Section B, with additional detail in activity-specific tables in Annex A. Chapter IV concludes the report by addressing the administration of the evaluation, including data management, dissemination, the roles and responsibilities of evaluation team members, and the timeline for evaluation activities.

Figure I.1. Overview of TAG activities and key outcomes

Activity	Objective	Budget (million USD)	Key outcomes
PAJI	Improve Kosovo's judicial sector through open data and efficiency improvement projects	4	• Increased public access to judicial data • Improvements in judicial efficiency
EDC	Improve access to real-time air quality data and disseminating the data to the public	3	• Increased availability of real-time and forecast air quality information • Increased public awareness of effect of air quality on health
KODC	Implement a series of open data challenges to involve civil society in TAG activities	1.3	• Increased engagement between GoK and civil society • Increased usage of open data for analyses and decision-making

II. Threshold Program overview

In this section, we provide an overview of the TAG project and details about the three relevant activities. We also discuss the theory of change and the causal pathways driving the program logic. In Section C, we provide a brief synthesis of literature relevant to TAG's three activities. We close each subsection of the literature review with the key contributions each activity-level evaluation will be able to make to the respective body of research.

A. Overview of the project and implementation plan

The objective of the TAG project is to improve judicial, environmental, and labor force data, making the data more accessible to the public to increase the use of data by civil society, business, and the government to support decision making. The three activities of the TAG project are the Public Access to Judicial Information (PAJI; US$4 million) Activity, the Environmental Data Collection (EDC; $3 million), and the Kosovo Open Data Challenge (KODC; $1.3 million). These three activities were designed to be synergistic, each contributing to the objective of establishing a financially sound, transparent, and accountable institutional basis for data delivery and economic expansion for Kosovan households, civil society actors, and firms. **Table II.1** summarizes the key components of each activity, which we describe in greater detail below.

Table II.1. TAG activities and their key components

Activity	Key components
PAJI 	• Creating an Online Data Platform (ODP) to enable public access to aggregate judicial data as well as data disaggregated by demographics and other relevant categories • Launching a Case Tracking Mechanism (CTM) that provides individual access to case information for authorized public users • Supporting Kosovo Judicial Council (KPC) communication, outreach, and publication
EDC 	• Installing network of air quality sensors and monitoring platforms • Launching Air Quality (AQ) monitoring and forecasting tools through websites and downloadable phone apps • Communications campaign to inform citizens about AQ and AQ data to promote behavior change
KODC 	• Assembling datasets in four DigData Challenges (Judicial, Air, Energy, Labor Force) for civil society organizations and companies to develop data innovations • Disbursing up to $1.3 million in total to DigData grantees to develop data products for government agencies and/or the public

Public Access to Judicial Information (PAJI) activity

The PAJI activity aims to increase public awareness of and access to judicial information, thereby enabling civil society to advocate for judicial reform more effectively. The activity builds on prior and ongoing judicial reform efforts by the Kosovo Judicial Council (KJC), Kosovo Prosecutorial Council (KPC), and the Ministry of Justice (MoJ) with support from the United States Agency for International Development (USAID) and the Norwegian Ministry of Foreign Affairs. PAJI consists of three sub-activities: (1) creating an Online Data Platform (ODP) for the public to access judicial data disaggregated by different demographic and other indicators; (2) developing a Case Tracking

Mechanism (CTM) to allow individual access to case information for authorized public users; and (3) supporting communications between judicial and legal institutions and to the public. Both ODP and CTM will draw data from the Case Management Information System (CMIS), which was developed with support from the Norwegian government to increase judicial efficiency by helping judges organize their extensive caseloads. The ODP and CTM aim to enable public access to judicial data, as well as help citizens access their individual case information online to increase judicial transparency. To the degree possible, the ODP also provides judicial data disaggregated by variables of interest to gender and social inclusion issues. Kosovo Legal Services Company (KLSC) in Consortium with B&S Europe were contracted to deliver the assessment and supervision of PAJI activity, whereas InfoSoft Systems sh.p.k, (in joint venture with Edusoft d.o.o., Nextsense Ltd., and Infosoft Systems sh.p.k. Albania) were contracted to implement hardware and software components of these sub-activities. The hardware and software implementation of PAJI activities was supposed to take place over a one-year period from 2021 to 2022 before the Kosovo Threshold Program end date of September 30, 2022. Following the end of the Threshold Program, KJC will be responsible for maintaining the CTM and ODP along with the trainings and engagement with public stakeholders.

Environmental Data Collection (EDC) activity

The Environmental Data Collection (EDC) activity aims to improve the quality and availability of air quality (AQ) data, which can inform the public of their health risks from pollution exposure, and support analyses by civil society and government on air quality in the country. To strengthen the quality of Kosovo's air quality data, the EDC activity seeks to improve the air quality monitoring network's hardware and software through the Kosovo Environmental Protection Agency (KEPA), the Kosovo Hydrometeorological Institute (KHMI), and the National Institute of Public Health (NIPH). Under this activity, TAG investments launched an online portal and a phone-based app that provides air quality forecasts and real-time information from the new monitoring equipment and new communication equipment for existing stations. To promote awareness and use of the data across relevant ministries, the Millennium Foundation Kosovo (MFK) conducted a needs assessment of data used by NIPH and KEPA. The activity also included an awareness campaign of the causes and health impacts of air pollution to encourage Kosovans to advocate for a cleaner environment, take steps to reduce their contribution to pollution, and minimize health risks from air pollution exposure. For example, on days of particularly poor air quality, the public was advised to either avoid strenuous physical activities outdoors, or at least minimize time spent engaging in such activities. These campaigns identified groups which are particularly vulnerable to air pollution, such as pregnant women, and developed dedicated messaging and outreach to target those groups. The activity's long-term goal was to foster greater awareness about air quality and its impacts, with the intent of spurring greater collaboration between civil society and the GoK on issues related to air quality.

As of June 2022, all components of the EDC activity had been fully implemented. The Air Quality portal was completed and operational in April 2021. The air quality outreach and behavior activities, including air quality info days, trainings, and media appearances concluded in 2021. The air quality forecasting services, including the website and related app, were completed in December 2020. Following the end of the Threshold Program, KHMI will be responsible for maintaining the air quality network, along with all software connections to websites and phone apps.

Kosovo Open Data Challenge (KODC) activity

The KODC activity aims to foster a more productive partnership among Kosovo's government, private sector, and civil society by supporting data-driven innovation and

promoting a culture of information sharing and evidence-based decision making. The activity sought to award up to $1 million in grants through a competitive process to individuals or organizations with innovative ideas about how to use, analyze, and present data to influence and support the Government's analytical and public communication needs. Grants were to be offered for either short-term data communication and outreach efforts to engage citizens and civil society or longer-term "proof of concept" projects for grantees to develop innovative data products for government institutions and their partners in the private sector and broader citizenry. Open data challenges, branded as "DigData Challenges," have supported grantees in four areas: (1) time use and labor force data, with an emphasis on analyzing gender-specific barriers through the DigData Labor Force challenge; (2) judicial data through the DigData Judicial Challenge; (3) environmental and energy data through DigData Air; and (4) DigData Energy.

In July 2018, MFK released a call for DigData Labor Challenge grantees and provided a two-month window for submission. After judging submissions, MFK awarded seven labor challenge grants in October 2018 and grant implementation began in November. After several delays, MFK released a call for applications for the DigData Air Quality Challenge in June 2019 and the Energy and Judicial Challenges in January 2021, giving applicants two-month windows to submit their materials. Eligibility criteria for prospective grantees in all challenges were highly inclusive, allowing any individual, group, or organization to apply.[1] Twenty-two applicants were accepted, with seven grantees in DigData Labor Force, four in DigData Air Quality, six in DigData Judicial, and five in DigData Energy. The goal of the activity was to engage, support, and connect local innovators—in civil society, the private sector, and academia—to relevant government entities, thus increasing the use of open data, producing tools, and performing data analysis to respond to government needs. In turn, relevant government entities would receive support from DigData grantees to improve data-sharing practices, co-implement innovative solutions to judicial, environmental, and energy questions, and implement the solutions identified through the KODC. The activity emphasized the importance of gender and social inclusion. For example, in the Labor Challenge, MFK encouraged grantees to focus on women's labor force issues and analysis, and across all challenges, MFK selected women-run organizations to pursue their projects. Successful grantees were slated to receive between $1,000 and $50,000 to support their activities. Grant agreements stipulated that after the challenges ended, GoK would receive royalty-free license to use and maintain the functionality of grantees' products, though the intellectual property rights would remain with the grantees. In cases where MFK and grantees determine that GoK does not have the capacity or interest to maintain the products, MFK planned to encourage grantees to maintain their products and seek additional funding to scale up their solutions after their grant with MFK ends.

B. Theory of change

The theory of change (TOC) illustrates how sub-activities of the TAG project contribute along a causal pathway to the overall goal of the Threshold Program (Figure II). It gives an overview of the key sub-activities, their anticipated outcomes, and how those outcomes work synergistically to support TAG's overall long-term outcome of increasing business investment in Kosovo to reduce poverty through economic growth.

The TOC posits that providing transparent data from diverse government agencies will improve public perceptions of government effectiveness and responsiveness. These data will help address real and perceived weakness in rule of law, and government accountability and transparency in Kosovo. The

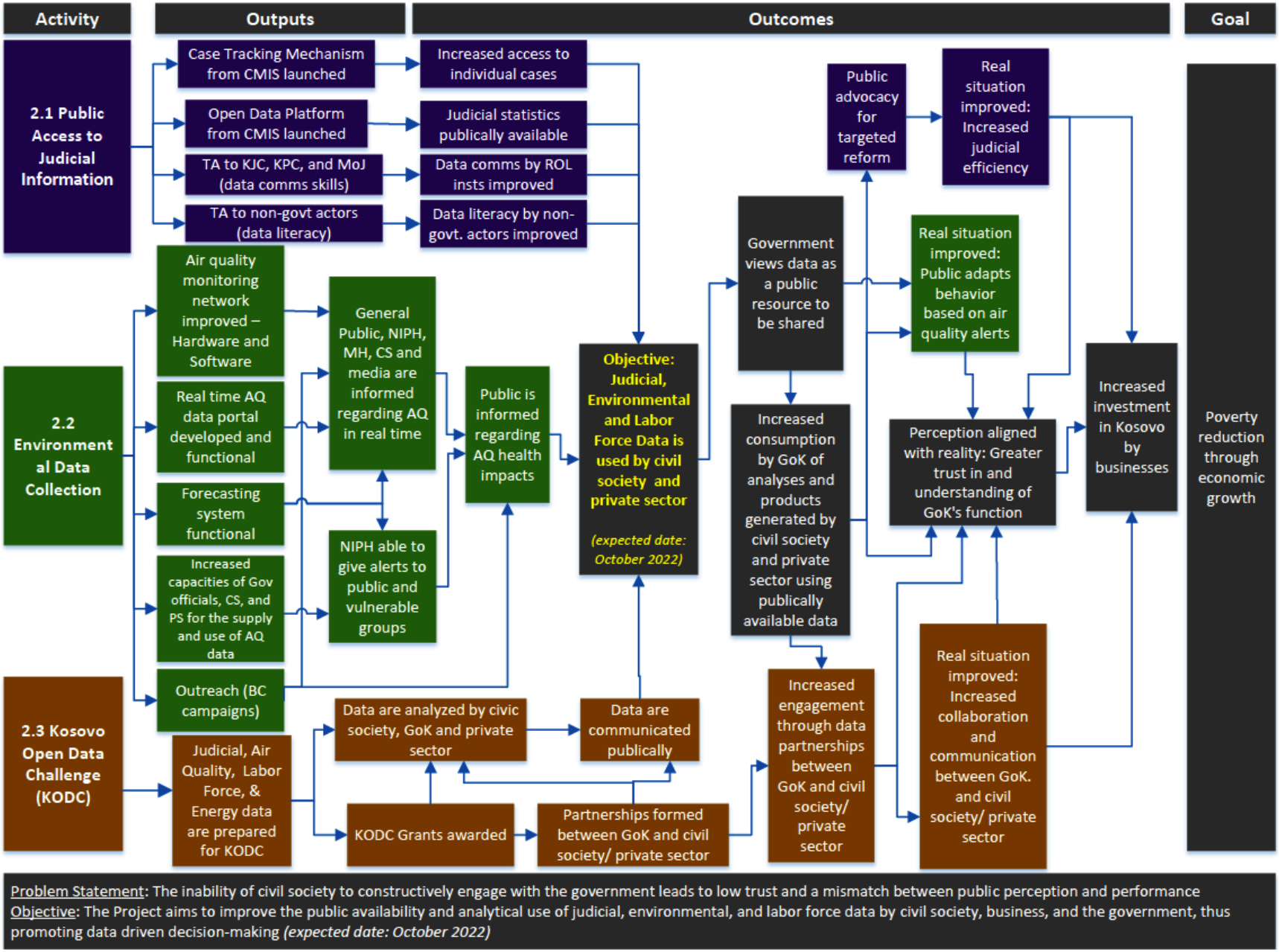
[1] Restrictions only excluded organizations that the World Bank or U.S. federal government have barred from procurement, or anyone prohibited from commercial relations with Kosovo.

current reality is that there is a disconnect between the laws in place mandating government transparency and the ability of civil society to constructively engage with the government. This has led to low public trust and incongruence between perceptions and performance, regardless of the government's actual performance.

Through complementary investments in open data, dissemination, and learning activities, the project theory anticipates that civil society and other beneficiaries will be able to access judicial, environmental, energy, and labor force data, thereby increasing trust in government and aligning public perception and trust in government institutions with reality. Access to air quality data in real time and access to personal case information or anonymized, disaggregated judicial data will allow individuals and the public and private sectors to interface with government information and potentially improve public perceptions of diverse government institutions. By offering access to labor force data and analysis, KODC DigData Challenge solutions could improve public understanding of unemployment, other labor dynamics, and government policy responses. Similarly, DigData products that draw on energy data could improve public understanding of Kosovo's energy consumption and mix, in turn stimulating citizen and civil society engagement on energy issues.

The TOC relies on a few key assumptions. First, that the usefulness of these open data sources will be clear and that civil society and the private sector have incentives and capacity to use the data for their purposes. As the public and private sectors use the data, the GoK will in turn view these data as a "public resource to be shared" and use the data-driven analysis and reports from civil society to inform government reforms and policies. This logic is in accordance with the KODC efforts to increase data partnerships between GoK and civil society through the exchange of data and subsequent data analyses. For this to happen, the TOC assumes the data will be presented in a way that is actionable and useful to project beneficiaries.

Figure II.1. TAG theory of change



Source: MCC 2022

C. Literature review

To provide context for our evaluation of the TAG project, we review the existing literature related to open data and on open data interventions on judicial transparency, air quality, and open data challenges. For each activity, we discuss gaps in the literature. We conclude with the policy-relevance of that this evaluation and how it will contribute to the existing literature.

1. Open data

Open government and inclusive policymaking practices can build public trust and support economic and democratic development (Matasick 2017). While transparency through open data may reveal that some trust in government is unwarranted—for example, by revealing corruption—evidence suggests it also provides a space for warranted trust to grow when combined with citizen engagement strategies.[2] In the UK, researchers found the government's criminal justice transparency program promoted warranted trust among the public, but open data practices were more effective when paired with deliberative processes (O'Hara 2012). Citizen satisfaction also plays a role in transparency and trust: researchers in the European Union (EU) used structural equation modeling to show that open data supports the trust of citizens across the EU in public institutions, with citizens' satisfaction playing a mediating role between the two variables. The effects of open data on trust are stronger among more educated, younger, and more digitally connected citizens (Gonzálvez-Gallego et al. 2019). Indeed, younger generations (Millennials [born between 1981 and 1996] and Generation Z [born in 1997 or later]) already tend to trust their public institutions more than older generations, and open government data further bolsters this trust (Gonzálvez-Gallego and Nieto-Torrejón 2021a). This is because younger people are more aware of the internet's role in providing public access to information and thus tend to value governments' efforts at open data transparency more highly. But data openness alone is not as powerful a driver of trust as the data coverage within—the disaggregation, historical availability, and administrative levels covered in the data (Gonzálvez-Gallego and Nieto-Torrejón 2021b). Thus, increasing accessibility to existing datasets is the first step toward transparency but does not automatically lead to a democratic government (Stojkov et al. 2016).

Open data policies and data portals to support these policies are common tools governments use to increase transparency. In Eastern Europe and the Western Balkans, governments have pursued open data policies not just for transparency and anti-corruption purposes, but also to improve efficiency, innovation, environmental awareness, and competitiveness. Kosovo has led in terms of civil society capacity building around open data and their focus on GSI in the space (Sadiku and Chung 2019). Open data portals are one channel to improve transparency (Lnenicka and Nikiforova 2021), but research from the Netherlands suggests such portals may not prompt macro-level change. Ruijer and Meijer (2020) gathered similar insights in living lab experiments with civil servants, students, and researchers. The researchers found that

[2] Even the perception of corruption prevents businesses from investing in a country's economy due to perceived economic and reputational risks (Cieslik and Goczek 2018; Zakharov 2019; Krasniqi 2013). Kosovo performed consistently worse in 2015 and 2016 than competitor countries on at least three corruption indicators in Transparency International's Corruption Perceptions Index (Transparency International 2016), though measurable metrics of corruption show these perceptions of corruption in Kosovo may be inflated (Krasniqi 2013). A biannual survey of Kosovans cites that, as recently as November 2021, 28 percent of respondents perceived corruption in courts, 19 percent in the central government, and 29 percent in the Tax Administration of Kosovo (UNDP and USAID 2021).

a multistage and iterative open data intervention[3] can produce changes in the data practices and culture of specific agencies but may not result in society-wide changes.

Researchers use several assessment methods to evaluate the quality of open data from multiple dimensions. Raça and coauthors (2021) developed and piloted a framework and application to assess open data portals of six Western Balkan countries. Using the Five Star assessment methodology introduced by Berners-Lee (2020), Raça and colleagues measured the availability and quality of datasets on national open data platforms, including Kosovo's opendata.rks-gov.net, and found that all six countries had average scores between 2 and 3 stars on the 5-star scale. The principal issue with the portals was the low-quality dataset formats (Raça et al. 2020). A dashboard providing real-time monitoring of open data portals can help policymakers and open data advocates identify areas for improvement in their data publications (Raça et al. 2021).

Other open data assessment methods include the Open Data Readiness Assessments (ODRA; World Bank 2015), the Global Open Data Index (GODI; Open Knowledge 2017), and Open Data Barometer (ODB; W3C 2017). Although the World Bank has released an ODRA for Serbia, it has not yet released one for Kosovo. The GODI, which crowdsources its analysis of global open data releases, monitors whether data are provided in a way that is truly accessible to citizens, the media, and civil society (Stojkov et al. 2016). Kosovo received a 26 percent openness score across all indicators according to their measures, highlighting their need for the open data interventions through TAG (Open Knowledge 2017). However, even where data are available and of high quality, there remain key differences in open data usership; Sadiku and Chung (2019) note that civic users in Kosovo and the Balkan region tend to use open data more visibly than the governments that publish it. The researchers suggested there are three main civic groups of open data users: (1) advocacy organizations and the media, (2) tech incubators, and (3) academics. The TAG KODC DigData Challenges engage all three civic groups. All three TAG activities seek to improve government use of open data and analyses conducted by outside stakeholders.

2. Judicial transparency

In the process of bidding for EU accession, Kosovo has established "increasingly sophisticated laws on paper" (Zogaj et al. 2017, p. 24) but has struggled to implement them (Pepaj 2015). The judiciary is understaffed and ill-equipped to handle the sheer volume of cases, with each judge handling about 800 pending cases per month (Pepaj 2015). Bajpai and Meyers (2020) found that the efficient operation of institutions in the judicial system is greatly undermined by capacity limitations and under-resourcing in terms of finances, human resources, and facilities, particularly in low-income nations. Despite hundreds of millions of dollars invested by donors, the judiciary in Kosovo cannot yet effectively provide redress for the business community against poor application of the law by the public administration because these investments have largely focused on criminal rather than civil or commercial issues (Rashiti 2019). Further, while the Constitution of Kosovo guarantees citizens the right to public documents, the justice system remains far detached from the public. Political trust is essential for democratic governance but is hobbled in Kosovo by citizen perceptions of corruption, challenges related to security and stability, and poor government capacity (Babamusta 2019).

[3] The intervention consisted of five phases, each informed by the last: (1) workshops to identify open data needs, (2) open data innovation development, (3) experimentation, (4) capacity building with relevant institutions, and (5) scaling to government agencies.

A diverse set of interventions, including online courts, audio-visual courtroom elements, case management (Dahmani and Vermeille 2017), and digitization of judicial records (USAID n.d.), have been implemented to address these issues to improve judicial efficiency. Some countries' and international organizations' e-justice systems, such as the EU's *e-curia*, have been operating for more than a decade and have allowed individuals greater access to their case data (Reiling and Contini 2022). Though there is limited empirical evidence on the subject, increased information and communications technology (ICT) in a judicial setting has been linked to positive outcomes and increased efficiencies in some cases (Raca et al 2021; Kucera and Chlapek 2014). Reiling (2020) posits that encoding procedural decisions and court work processes into the digital court environment may affect fair procedure and judicial impartiality. Fedushko et al. (2019) found that digitizing judicial services sped up workflows, cut down on time spent on jobs that can be automated, reduced bureaucracy, lowered the risk of corruption, and provided other relevant advantages. Case tracking mechanisms (CTMs) are common in many sectors—such as in law for managing client cases and in medicine for managing patient cases—but are not systematically used across national judicial systems (Lee et al. 2020; Reuters 2022). CTMs can summarize essential information in a timely way, such as the International Patient Summary used to track COVID-19 patients across borders in an effort to quickly track the virus (Lee et al. 2020). According to USAID's Case Tracking and Management Guide, CTMs can give judges a complete record of cases and help them with case-control and decision-making (USAID n.d.). CTMs can also help judges create schedules and effectively manage their caseloads. The CTM implemented through PAJI may help reform Kosovo's judiciary as an e-judiciary that uses a central database environment and synchronously uses ICT infrastructure to exchange data and documents electronically inside the judicial system and with relevant external ICT systems through the Case Management Information System environment (IMG 2013).

While advancing ICT in the judicial system might seem like an obvious step forward, human control is essential in all stages of the judicial process to safeguard fair procedures (CEPEJ 2018). ICT systems have been implemented in many places, but some have reverted back to earlier manual systems (Reiling and Contini 2022). Barriers to ICT progress and sustainability include security concerns (USAID n.d.), burdensome time demands on judges and staff (Velicogna et al. 2011), and other logistical issues like commissioning and maintaining the appropriate software (Fedushko 2019). Stojkov et al.'s (2016) open data assessment of the Western Balkans found that, when the countries assessed did have access to open data, it was spread out between specialized websites and not available in an accessible way.

3. Air quality

Ambient air pollution is a leading cause of premature death globally and ranks among the top 15 risk factors in Eastern Europe (Lim et al. 2013, GBD 2019 Risk Factors Collaborators 2020). The morbidity effects of air pollution exposure extend beyond respiratory issues (Janke 2014), and have been also linked to cardiovascular disease, central nervous system dysfunction, and cancer (Manisalidis et al. 2020). While public health research has extensively documented a spectrum of negative health impacts, exposure to harmful levels of ambient air pollution adversely affects outcomes ranging from academic performance and cognitive functioning (Lavy et al. 2014; Chen et al. 2018; Zhang et al. 2018; Graff Zivin and Neidell 2018; Carneiro et al. 2021), labor productivity (Fu et al. 2017; Aguilar-Gomez et al. 2022), mental health (Chen et al. 2018), to economic activity more broadly (Dechezleprêtre et al. 2019).

In Kosovo, exposure to both ambient and indoor air pollution is also a leading cause of morbidity and premature death. The World Bank (2019) estimates that annual mortality costs from pollution exposure are 2.4 to 4.7 percent of the country's GDP. An estimated 760 people die prematurely each year because of air pollution in Kosovo, whose capital city routinely experiences pollution concentrations comparable

to the most polluted cities in China and India (World Bank 2019). While coal-fired power generation is a significant source of pollution, it is not necessarily the dominant source. Small combustion accounts for the majority of both PM_{10} and $PM_{2.5}$ emissions (World Bank 2019), while industry is the key source of NO_x and SO_2 (AMMK 2020; NIRAS and Atmoterm 2020). In Pristina, residential heating is a key factor for spikes in $PM_{2.5}$ readings during winter nights and gives the city a highly seasonal pollution profile (Edwards 2020).

Regulatory approaches intended to stimulate pollution abatement or create markets for tradable emissions permits have a proven record of reducing pollution in the long run (Shapiro and Walker 2018), but they provide no immediate relief against acute episodes of poor air quality. Compounding this problem is that people may not know which risk reduction actions they should undertake (Weng et al. 2021). One prevalent method of supporting short-run pollution avoidance is providing information about pollution levels. This information can be communicated through a range of channels including government websites (Naiker et al. 2012), TV and radio broadcasts (Limaye et al. 2018), newspaper articles (Ramondt and Ramírez 2020), SMS air quality alerts (Saberian et al. 2017; Hanna et al. 2021), physical signage in public places (Riley et al. 2021), and phone apps (Delmas and Kohli 2019). Air quality information is often bundled with public health recommendations based on the prevailing pollution concentration, such as postponing vigorous outdoor exercise, wearing a mask outdoors, and running air purifiers indoors with windows and doors closed. The accuracy and usefulness of air quality information depend on the proximity, density, and maintenance of air quality monitoring stations. Yet the rise of inexpensive sensors that private citizens can install, such as PurpleAir, has broadened the reach of government sensor networks and facilitated finer-resolution air quality modeling (Bi et al. 2020) than is possible with regulatory monitors alone. Although readings from such sensors may be systematically biased, they can still be useful by applying calibration factors and merging with conventional monitoring networks (Barkjohn et al. 2021; Bi et al. 2020).

Few studies have estimated the impact of supplying air quality information on behavioral or health outcomes. Barwick et al. (2019) offer some of the most compelling evidence for a range of outcome domains in their analysis of China's staggered introduction of an air quality information system over a two-year period. They find that merely providing air quality information—uncoupled from regulatory change or policy reforms—affected retail patterns, purchases of indoor air filters, housing market corrections near major polluters, and a reduction in mortality response that could be explained through a pollution avoidance channel. Likely critical to these findings is the high baseline pollution concentrations in China—cities affected by this rollout were among the most polluted cities in the world. Areas with lower pollution levels might not experience the same changes in behavior and health outcomes. Other examples of causally identified impacts of air quality information or alerts tend to be more narrowly focused, such as on outdoor bicycling trips (Saberian et al. 2017) or whether individuals close their windows (Hanna et al. 2021).

Air quality alerts and systems could be an effective means of promoting public health, since they make visible the threat of something invisible, but human psychology is a potent inhibitor and undermines the effectiveness of such information. Even though people cannot accurately perceive actual pollution levels via their senses (Semenza et al. 2008; Cori et al. 2020; Boso et al. 2022), some studies find that personal perceptions, not objective indicators, are the primary driver of risk reducing behavioral changes (Semenza et al. 2008; Hanna et al. 2021). Even when individuals use air quality alerts instead of their own perceptions to inform behavior like whether to exercise outdoors, psychological factors still influence whether individuals translate information to effectively reduce health risk. For example, AQ app usage on

phones tends to drop off precipitously within months of downloading the app, instead of becoming a long-term habit (Delmas and Kohli 2019). If people stop using the app, then their actions will not be based on actual AQ conditions. If people are responding to deviations in air quality levels, and not to the levels themselves (Gurajala et al. 2019), they may be more regularly exposed to poor air than is advisable. Even if people are receiving accurate information about air quality, they may believe it does not adversely affect them and should only prompt others to change their behavior (Bickerstaff and Walker 2001). Recognizing the psychological factors that lead people to engage in avoidance behavior when pollution levels are acceptable, but not engage in those behaviors when air quality actually is poor, should be a central element in improving the design and effectiveness of air quality information systems (Oltra and Sala 2015).

4. Open data challenges

Open data challenges are a common way for public institutions to engage the private sector and civil society in addressing social, economic, and environmental issues while increasing government transparency. Governments or foundations tend to initiate open data challenges by identifying policy issues and relevant datasets, framing the competition by defining the issues and opportunity areas, and releasing a call for applications with eligibility information. The Nesta Challenge Series in the UK (Nesta 2022) provides an example of a large-scale, foundation-funded open data challenge, with competitions for companies and civil society organizations (CSOs) in jobs, food, culture, and other areas. Competing companies or CSOs submit proposals to develop products or services that use pre-selected or other available open datasets, and the challenge administrators, often with outside experts, judge the quality and value of the proposed products and services. In the United States, the data.org Inclusive Growth & Recovery Challenge, funded by the Mastercard and Rockefeller Foundations, awarded prizes to companies and organizations working on data-driven solutions to challenges in the themes of cities and towns, access to capital, and jobs of tomorrow (Data.org 2022). Similarly, the Open Data for Good Grand Challenge, funded by the U.S. Census Bureau and delivered by The Opportunity Project, provides monetary and in-kind awards to private sector and civil society-based data teams that have produced digital tools that solve public problems (U.S. Census Bureau 2022). Challenge administrators grant funds to companies and CSOs with proposals to develop high-quality, high-value products or services or scale up those that already have prototypes. Because the grants are time-bound, they often stipulate that grantee products and services, such as apps, are transferred to beneficiary institutions, such as government agencies, to maintain them after the challenge closes.

There is little evidence on the effectiveness of open data challenges in addressing social, economic, and environmental issues. Initial reports on the Data.org project suggest the challenge attracted high-capacity organizations with innovative approaches to data science, including machine learning and market analysis to support renewable energy development among women entrepreneurs, but an assessment of the results of their work is still forthcoming (Data.org 2022).

In the Balkans, a growing culture of open data provides a rich landscape for open data challenges. For example, Kosovan programs that pair open data with ICT skills training for women and girls (such as Open Data Kosovo and Girls Code Kosova) can help them overcome gender-related barriers to employment and civic participation by fostering empowerment in technological, social, psychological, and political realms (Domagala 2020). Hackathons in Serbia and Georgia show promise in terms of linking technologists and advocacy organizations to explore data applications (Sadiku and Chung 2019). Moldova and Ukraine are also building platforms for e-participation, which can improve citizens' engagement in democratic processes and allow them to impact public policy more easily (Khutkyy 2019).

In Ukraine, government agencies collaborated with CSOs to build the Coalition for the Advancement of E-Democracy, which mapped civil society engagement steps and oversaw development of e-participation tools, such as e-appeals (individual suggestions or complaints to the government), e-petitions (collective e-appeals), procurement data platforms, and other tools. Based on web traffic, public uptake of these tools is high, and qualitative data suggest they improve transparency and accountability between citizens and the government.

5. Policy relevance

Overall, the literature on open data strategies is wide in scope, but there is little empirical evidence of their impacts. Many studies focus on challenges and innovations with open-data implementation, yet few studies fully trace the effects of those factors on public trust and collaboration between civil society and government. Our evaluation will offer a more robust assessment of TAG's open data contributions by combining quantitative evidence on the changes associated with the project with a comprehensive qualitative analysis to understand how and why these changes are observed. The qualitative interviews with stakeholders involved in PAJI, KODC, and EDC will give insight into the causal mechanisms and impacts of the interventions themselves.

PAJI. The proposed performance evaluation of PAJI will provide useful evidence on the effects of technical assistance in judicial settings. This evaluation will add to the overall evidence base on judicial reforms and will provide guidance to donors, policymakers, and practitioners implementing or improving e-justice systems (particularly in the Balkans). The evaluation will also generate insights for MCC on the challenges implementers and stakeholders face when maintaining judicial processes and operations while overseeing major technical reforms, capacity building, and maintenance of new data infrastructure. We also anticipate our findings will indicate how open data systems can be best maintained after development funding ceases (after the end of the Threshold Program). Finally, our evaluation may be able to contribute to the evidence on how judicial transparency policies could influence public understanding and trust in judicial systems and the efficiency of those systems.

EDC. The existing literature primarily focuses on the linkage between air quality information and individuals and under-emphasizes the role of other actors. The EDC evaluation will contribute to the literature by prioritizing CSOs and untangling their role as a mediator between government and the people in engaging and responding to air quality data. The evaluation will invert the conventional directionality of democracy and institutional capacity affecting environmental conditions (Bernauer and Koubi 2009, You et al. 2015) and instead examine how environmental conditions influence governance and public trust in officials. We are unaware of any research that investigates how the provision of air quality information influences the dynamics and collaborative quality of relationships between civil society and the government, which will be a genuine contribution from this evaluation. The evaluation will also contribute to a relatively small literature examining how the public does or does not engage with high-quality air quality monitoring network data.

KODC. Our evaluation of the KODC activity will contribute to the evidence on open data challenges in three ways. First, it will provide insights into how grant competitions can stimulate data-related engagement between civil society/private sector and government actors. Second, it will develop lessons on what types of data products and innovations are most valued by beneficiaries (whether they are citizens or government institutions) in the Kosovan context. Third, we will explore whether increased collaboration and communication between civil society/private sector and the government are sustained

beyond the life of the challenge grants, and the implications of that sustainability in terms of improving an open data culture.

III. Evaluation Design

In this chapter, we describe our proposed mixed-methods performance evaluation design of the TAG investments. Our evaluation is structured as four distinct components, that consist of three activity-specific evaluations along with one cross-cutting evaluation to examine synergies among the activities and other dimensions of TAG that extend beyond the scope of any individual activity. We present the research questions for each component in Section A. To systematically determine whether these questions can be properly answered given the available documentation and data, as well as critically appraise the project's TOC, we conducted an evaluability assessment. We summarize the results from that assessment in Section B and include the complete assessment findings in Appendix B. We describe the research methods to be used in the four evaluations in Section C, and we outline the data sources they will draw on in Section D. In Section E, we lay out the specific evaluation approaches for the activities and the cross-cutting evaluation. A brief discussion of key evaluation risks and risk-mitigation strategies closes out the chapter in Section F.

A. Evaluation questions

Our evaluation will answer the activity-differentiated research questions listed in Table III.1.[4] In the sections below, we elaborate on the specific research methods we will use to answer these questions and the data sources they will draw on.

Table III.1. Research questions by activity

Activity	Research questions
PAJI 	**1.** Was the activity implemented according to plan (in terms of quantity and quality of outputs)? **2.** Did the activity achieve its targeted outcomes, particularly its stated objective, in the timeframe and magnitude expected? Why or why not? **3a.** Is there any increase in the GoK's use of analyses done by non-government entities, both analyses supported by MFK and those in general? Is there any evidence these types of analyses can factor into any policy decision making? **3b.** Did publishing judicial data through relevant government websites (particularly KJC) result in increased analysis done by NGOs? **3c.** Did PAJI result in increased engagement between government and civil society/media? **4.** How has the project contributed to citizens' use of judicial data in advocating for change? **5.** Did PAJI contribute to increasing trust and understanding of the judiciary system's functions?

[4] The numbering convention used in this report follows the convention used in MCC's original list of TAG evaluation research questions. Appendix A documents the differences between the research question text in MCC's request for proposals and the text we propose in our evaluation.

Activity	Research questions
EDC 	**1.** Was the activity implemented according to plan (in terms of quantity and quality of outputs)? **2.** Did the activity achieve its targeted outcomes, particularly its stated objective, in the timeframe and magnitude expected? Why or why not? **3a.** Is there any increase in the GoK's use of analyses done by non-government entities, both analyses supported by MFK and those in general? Is there any evidence these types of analyses can factor into any policy decision making? **3b.** Did the dissemination of air quality data through government websites affect activities by NGOs/CSOs, and if so, why? **3c.** Did EDC result in increased engagement between government and civil society/media? **4.** How has the project contributed to citizens' use of environmental data in advocating for change? **5.** Does the existence of transparent, government-produced air quality data, health advisories, and a national outreach and behavior change campaign create enough incentive for civilians to change their behavior (e.g., take actions to reduce the negative health impacts of air pollution)? If evidence of changed behaviors exists, who is adapting, and how and why have they adapted? **6a.** Does the existence of transparent, government-produced air quality data reduce the adversarial relationship between civil society and the GoK? If yes, whose attitudes and behaviors are likely contributors to these reductions? **6b.** Has inter-ministerial communication changed, e.g., between Kosovo Environmental Protection Agency (KEPA) and National Institute of Public Health (NIPH), and if so, why and how? **6c.** Are air pollution data available on a continuously updated basis? How accurate are the air pollution forecasts provided through the NIPH portal? What percentage of time does air quality exceed given thresholds? **7.** Did EDC contribute to increasing trust and understanding of government's function?
KODC 	**1.** Was the activity implemented according to plan (in terms of quantity and quality of outputs)? **2.** Did the activity achieve its targeted outcomes, particularly its stated objective, in the timeframe and magnitude expected? Why or why not? **3a.** Is there any increase in the GoK's use of analyses done by non-government entities, both analyses supported by MFK and those in general? Is there any evidence these types of analyses can factor into any policy decision making? **3b.** Did publishing energy, labor force, air quality, and judicial data through relevant government websites result in increased analysis done by NGOs? **3c.** Did KODC result in increased engagement between government and civil society/media? **4.** How has the project contributed to citizens' use of open data in advocating for change? **5.** Did KODC contribute to increasing trust and understanding of government's function?
Cross-cutting 	**1.** Did the program achieve its targeted outcomes, particularly its stated objective, in the timeframe and magnitude expected? Why or why not? **2.** Do the results of the program justify the allocation of resources for it? **3a.** Is there any increase in the GoK's use of analyses done by non-government entities, both analyses supported by MFK and those in general? Is there any evidence these types of analyses can factor into any policy decision making? **3b.** Has engagement between government and civil society/media increased? **4.** Is there a change in government employees' perceptions of government data as a public good or as a resource to be shared? If yes, how are government employees sharing data with the public (open data, website, reports)? If no, why?

B. Evaluability assessment

This section uses the five dimensions of evaluability in MCC's Project Evaluability Assessment Tool (MCC 2020, 2021) to assess the program logic and key assumptions of the PAJI, EDC, and KODC activities. These dimensions also allow us to document the activities' intended beneficiaries and assess their current monitoring and evaluation (M&E) plans. Where relevant, we highlight the implications of our assessment for the planned evaluations. We used various data sources for this assessment, including programmatic documents, discussions with stakeholders, and relevant literature. The documents include (1) M&E plans, logic models, and indicator tracking tables; (2) monthly, quarterly, and annual implementer reports; (3) meeting notes and slides; (4) implementer terms of reference; (5) analytical products, such as data visualizations on time use in the labor force; (6) product guides or manuals; (7) budget documents; (8) grantee concept notes and milestone reports; and (9) the constraints analysis. Table III.2 lists each dimension with its central question for assessing evaluability.

Table III.2. Evaluability assessment dimensions

Dimension	Overall question
1. Problem diagnostic	Is the problem clearly defined and is there sufficient evidence to support the problem diagnostic?
2. Project objectives and logic	Are the project objectives and theory of change/logic clearly defined?
3. Risks and assumptions	Are the risks and assumptions clearly defined with potential risk mitigation strategies?
4. Project participants and beneficiaries	Are project participants clearly defined and justified in terms of geographic scope and eligibility criteria?
5. Accountability and learning metrics	Are the metrics for measuring results for both accountability and learning clearly defined?

MCC provides up to nine sub-questions for each dimension to aid in the evaluability assessment. We applied these sub-questions to information available on each TAG activity to develop our understanding of evaluability. Table III.3 summarizes findings for each dimension, with evaluability coded as *strong*, *adequate*, or *of concern*.[5] We define an activity's evaluability in each dimension as "strong" if there is robust evidence and thorough planning to support the project's approach in that dimension, "adequate" if there is moderate evidence and planning to support the approach, and "of concern" if there is little or no evidence and little documented planning.

Table III.3. Summary of evaluability of TAG activities

Dimension	PAJI	EDC	KODC
1. Problem diagnostic	Adequate	Strong	Adequate
2. Project objectives and logic	Of concern	Adequate	Adequate
3. Risks and assumptions	Of concern	Strong	Of concern
4. Project participants and beneficiaries	Strong	Adequate	Adequate
5. Accountability and learning metrics	Adequate	Adequate	Of concern

Source: Mathematica evaluation team.

[5] Annex B provides all evaluability dimension sub-questions and our detailed assessment of each activity's evaluability.

Cross-cutting evaluability. Across PAJI, EDC, and KODC, the **project logic** that connects activity-level outputs to activity- and project-wide outcomes (and that connects those outcomes to the project objective) could benefit from additional evidence and justification. For example, the logic model shows that the outcomes of (1) *government views data as a public resource to be shared* and (2) *increased consumption by GoK of analyses and products generated by civil society and private sector using publicly available data* will produce the subsequent outcome of (3) *real situation improved: public adapts behavior based on air quality alerts*. However, the documents, stakeholders, and literature available do not clearly explain the assumptions and conditions that must be met for that logical link to hold. Particularly for outcomes where behavioral change is expected, we should also expect to see changes in the capabilities, opportunities, and/or motivations of actors (Koleros et al. 2020). In the example above, the TOC would suggest that, as a result of changes in government perspectives on data and increased GoK consumption of external analyses, the public has improved capability to engage with data, opportunities to apply the data to their lives, and motivation to care enough to make behavioral changes. These anticipated links are not fully examined in project documents, and there are also inadequate assumptions across several critical connections where activity-level interventions feed into larger project impacts. For example, the project TOC assumes that *increased collaboration and communication between GoK and civil society/private sector* will contribute to *increased investment in Kosovo by businesses*, but how, why, and whether businesses will act on noticeable changes in that collaboration and communication is somewhat unclear. However, one PAJI-related assumption outlined by MFK staff suggests that anticipated changes in investors' capabilities and opportunities may change their behaviors; specifically, increased transparency in the process and timeline for business-related court cases could stimulate interested investors to trust the local business environment more and use available judicial data and CSO reports for their decision making.

Several outcome indicators at the activity level also pose measurement challenges. For outcomes such as *Public adapts behavior based on the air quality alerts* (EDC), the evaluation team found there is not adequate evidence that the activity's interventions will change behavior at a level that can be measured with quantitative indicators and measurement strategies. Box III.1 provides insights on M&E indicators and subsequent subsections detail evaluability by activity. Section E in this chapter details our approaches for capturing changes in higher-level outcomes for each activity.

Box III.1. Insights on TAG M&E indicators and data source quality

Data Quality Review

MFK contracted IDRA Research & Consulting, a Tirana-based firm, to conduct a Data Quality Review (DQR) of the latest version of the Kosovo Threshold Program M&E Plan (Version 4). IDRA reviewed RELP and TAG indicators using the criteria that they should be direct, unambiguous, adequate, practical, and useful, and reviewed data sources using the criteria of validity, reliability, timeliness, consistency, precision, and objectivity/integrity. Using a 1 to 5 scale to indicate the degree to which criteria were met, with 5 representing the highest marks, IDRA scored all 14 indicators for TAG. PAJI, EDC, and KODC all had average scores below 3.5, indicating that most indicators failed to fully meet the indicator and data source criteria for quality (IDRA 2022). These findings aligned with our assessment of indicators and data sources proposed in the TAG M&E Plan and used in the indicator tracking table (ITT). IDRA provided clear recommendations for improving each TAG indicator, which can be summarized in four main areas.

1. Several TAG outcome indicators should be replaced with more measurable output-level indicators.
2. Indicator descriptions should be more concrete and measurement methodologies should be revised.
3. Develop new measurements for behavior change indicators and remove indicators that are not measurable as proposed.
4. Align monitoring data collection timetables with implementers' reporting periods.

MFK is seeking to adjust the indicators and data sources where feasible during the limited extension of the Threshold Program. For our implementation study of TAG, we will supplement weaker indicators provided in the final ITT with qualitative data collection and will use administrative data from beneficiary institutions. For the outcome indicators such as changes in perceptions and behavior, we will use qualitative data and analysis to draw insights related to the concepts behind the indicators. We will also draw on any indicators that MFK can revise and improve in the remainder of the Threshold Program. ◢

1. PAJI evaluability

Our evaluability assessment of the judicial activity found that several dimensions of the intervention appear to lack complete justification and planning.

Table III.4. Evaluability of PAJI

Dimension score	Summary of evaluability assessment
Adequate	**1.** The **problem diagnostic** of PAJI—that weak real and perceived rule of law constrains economic growth—is supported by evidence and clearly framed by the project in the local institutional context and political economy and in relation to ongoing interventions.
Of concern	**2.** While the PAJI **logic and objectives** are straightforward, the activity will not address all root causes of weak real and perceived rule of law (such as corruption or inadequate funding for the judiciary), suggesting that changes in outcomes further along the theory of change may not be noticeably affected. Finally, the sustainability of the activity's products, once transferred to government institutions, is not guaranteed.
Of concern	**3.** Project documents did not adequately interrogate **risks and assumptions**, which likely contributed to the delayed timeline of the activity's implementation. Critically, the causal link assumed between higher levels of judicial data use by citizens and civil society and improved trust in government is not adequately explored or supported by evidence.
Strong	**4.** Documents show the intended project **beneficiaries and participants** are judicial authorities, litigants, and the public. PAJI stakeholders indicated improvements in case tracking mechanisms may help a more specific group of beneficiaries: ethnic minorities who historically encountered barriers in fully accessing Kosovo's legal systems.[a]
Adequate	**5.** Finally, tools for **monitoring project implementation** and **assessing project results** are varied in quality and completeness (see also Box III.1). Most monitoring indicators examine changes at the activity- and initial-output levels (such as whether the ODP and CTM were launched or the number of government and CSO staff trained on new systems), and some indicators and questions to assess outcomes are vague.

Source: Mathematica evaluation team.

[a] Our evaluation will examine ways in which PAJI affected the legal experience of ethnic minorities through interviews with staff from KJC and CSOs working on rule of law issues.

These findings have implications for how the evaluation can examine PAJI's performance and impacts on outcomes of interest. To support the study, the evaluation team will gather additional information on the activity's logic, objectives, risks, and assumptions, including alterations to those areas (which will help us address evaluability issues in dimension 1). We will also seek additional information on how indicators were selected, on opportunities for data disaggregation (dimension 5), and on the timeline for the concurrent USAID intervention to digitize cases. Finally, we will trace the links between results stages in the PAJI TOC to identify whether and where assumed connections fail, and we will use that information to better understand the contribution of PAJI on changes observed in later outcomes (dimensions 2 and 3). A key factor for contextualizing PAJI sustainability is that the CTM relies on data drawn from case management information system (CMIS), which precedes PAJI activities and must be regularly maintained and have data input by judicial figures in order to feed timely and accurate data into CTM. Because the sustainability of desired PAJI impacts generated (such as increased transparency and efficiency) depends on the function of other systems, we will also assess the sustainability of those systems.

These findings have implications for how our evaluation can examine EDC's performance and impacts on outcomes of interest. To support the study, the evaluation team will gather additional information on how citizens found and used air quality information before EDC, how the project sees air quality data use affecting prospects for economic development (dimension 2), and how the activity selected applied GSI approaches and selected outreach locations (dimension 4). We will also supplement project monitoring and evaluation tools (dimension 5), and substitute originally proposed data sources that may provide a biased view of EDC's impacts—such as a stakeholder survey and the UNDP Public Pulse Survey—with app and website traffic data, Google trends data, and qualitative data (Section D of this chapter). These data sources will support our contribution analysis of EDC, which we describe in Sections C and E.

2. EDC evaluability

Our evaluability assessment of the environmental data activity found that it generally had adequate justification and planning for its interventions.

Table III.5. Evaluability of EDC

Dimension score	Summary of evaluability assessment
Strong	**1.** Robust evidence supports the **problem definition and diagnostic** that poor air quality leads to health issues, and likely also deters investment in Kosovo, posing an obstacle to economic growth. While the activity does not contribute to reducing air pollution, such as shifting away from fossil fuel combustion and current sources of domestic winter heating, it does address transparency in the exposure to environmental harms that people face. However, the diagnostic would be stronger if it had (a) comprehensive information on how citizens found air quality information before EDC and (b) stronger evidence that the lack of transparency about air quality was in fact a key impediment to air quality improvement.
Adequate	**2.** The EDC **objectives and logic model** are straightforward and provide clear links across activities to the efforts and outcomes of KODC. However, the anticipated connection between improved availability and use of air quality data and accelerated economic growth is not fully explained and the sustainability of the activity's effects on citizen/CSO air quality data use and advocacy is not clear. While specific government bodies have been designated to maintain the AQ network equipment and manage AQ information campaigns, the degree to which those bodies have the technical capacity and resources to carry out those tasks at scale is in question. The activity did not identify exactly how it will generate sufficient changes in the salience of air quality concerns or in beneficiaries' capabilities, opportunities, and motivations to drive behavior change, nor how it will reach enough people to produce measurable behavior change at the population level.
Strong	**3.** The activity clearly identified most **risks and assumptions**, as well as mitigation strategies, before implementation, and successfully applied those strategies to challenges such as the COVID-19 pandemic.
Adequate	**4.** EDC targets a clear (if broad) and justified set of **beneficiaries and participants**. However, the activity only appeared to consider gender and social inclusion (GSI) in the outreach and behavior change sub-activity and the criteria for targeting specific geographic areas for outreach were not fully articulated.
Adequate	**5.** EDC tools for **monitoring implementation** are strong, but indicators, data sources, and the timeline for **assessing results** are not as well documented or justified (see also Box III.1). Not all indicator targets were clearly justified in the documents or by stakeholders.

Source: Mathematica evaluation team.

3. KODC evaluability

Our evaluability assessment of the open data challenge activity found that several dimensions of the intervention appear to lack complete justification and planning.

These findings have implications for how the evaluation can examine KODC's performance and impacts on outcomes of interest. To support our evaluation, we will gather additional information on the specific issues that grantees aimed to solve and the expected connections between KODC outcomes and broader project outcomes (dimensions 1 and 2). We will also gather information on activity and grantee timelines, sustainability plans, risk and mitigation strategies, and project beneficiaries (dimensions 3 and 4). Finally, we will define, assess, and disaggregate outcomes wherever possible using qualitative data and quantitative grantee data, and will draw extensively from newly-available grantee milestone reports (dimension 5). These pieces of information will strengthen our analysis of the implementation of KODC and allow us to trace the links between expected results of KODC to identify whether and where projects assumptions fail. Additional information will also allow us to assess the contribution of KODC to improvements in transparency, open data collaboration, and stakeholder capacity and interests to support those outcomes.

Table III.6. Evaluability of KODC

Dimension score	Summary of evaluability assessment
Adequate	**1.** KODC documents and stakeholders clearly **define and diagnose the problem**: a lack of publicly available data—and lack of funds—for organizations and individuals to generate insights and advocate for improvements hampers trust in government and economic growth. However, grantees' subprojects do not use quantitative evidence of the problems they aim to resolve to justify their approaches, and it is not clear how the changes in information availability and data-driven advocacy will be sustained after the project.
Adequate	**2.** The KODC **objective and logic model** are clear and align with the problem diagnostic and the PAJI and EDC activities. However, the timelines for DigData Challenges (and those of grantees) vary, and the connections linking grantees' projects to sustained changes in data-driven decision making and economic development are not clear.
Of concern	**3.** TAG-wide materials define **risks and mitigation strategies** clearly, but KODC and grantee-level information does not adequately detail risks, assumptions, program monitoring for blind spots, or risk mitigation strategies.
Adequate	**4.** The Threshold Program M&E plan identifies and justifies target **participants and beneficiaries** for each activity, and KODC provides clear eligibility and selection criteria for grantees for each DigData Challenge. However, grant manuals do not clearly lay out how grantees themselves should target participants or a geographic distribution of work, and grantees provide variable amounts of information on audience and participant selection.
Of concern	**5.** KODC tools for **monitoring project implementation** and **assessing project results** are varied in quality and completeness (see also Box III.1). While activity-level monitoring is available for KODC, many annual indicator targets in the ITT are missing, high-level outcomes lack targets, and none of the KODC indicators in the ITT are disaggregated by gender, age, or income. Grantees provide monitoring data in milestone reports, but not all provide full or consistent data disaggregation.

Source: Mathematica evaluation team.

C. Research methods

Our evaluation approach consists of complementary research methods that are well-suited to answering the research questions in Table III.1. We will draw upon methods that systematically examine how TAG activities were carried out, along with methods to appraise the degree to which behaviors, perceptions, and the policy environment have been influenced by some component of TAG programming, or by the project (Table III.7). We describe our proposed research methods in this subsection, drawing attention to each method's core strengths and the specific capacities in which we plan to use them. In Section III.E we elaborate which methods will be used to evaluate which activities, since each activity has a distinct set of research questions.

All of the activity-level evaluations will include an **implementation analysis**, through which we will document how activity components were carried out, highlight when implementation materially deviated from plans (along with the presumptive causes for the deviation), and note which lessons and recommendations emerged from TAG implementation. Using the project's M&E plan (Millennium Foundation Kosovo 2021) as a guidepost, we will compare the actual outputs against the M&E targets and follow up with relevant stakeholders to understand the cause of any observed gaps. This analysis is based on primary qualitative data (key informant interviews [KIIs] with MFK staff, implementers, and government agencies that were involved with the project during the implementation phase) and secondary qualitative data (implementers' progress and annual reports, meeting minutes, contracts, and terms of

reference). In order to develop a coherent narrative about the project's implementation, we will triangulate all of the evidence from different sources and identify any differences among those sources' data or explanations. Where appropriate, our implementation analysis will draw upon additional methods described below that offer systematic techniques for drawing conclusions and interpreting the available evidence.

Table III.7. Research methods used by evaluation

Research method	Evaluation			
	PAJI	EDC	KODC	Cross-cutting
Implementation analysis	✓	✓	✓	✓
Contribution analysis with process tracing	✓	✓	✓	✓
Political economy analysis	✓		✓	
Descriptive trends analysis	✓	✓	✓	✓
Qualitative thematic analysis and triangulation	✓	✓	✓	✓
Correlation analysis		✓		

Note: Contribution analysis with process tracing is an overarching approach that connects other research methods in identifying the contribution of interventions to changes observed in outcomes where other interventions may also have an influence.

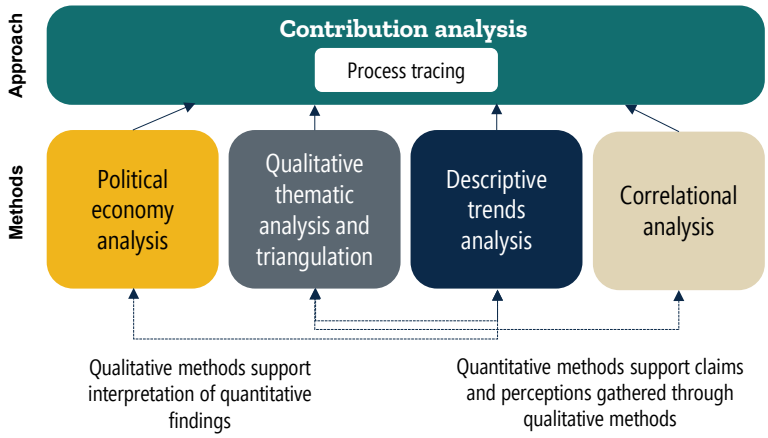
Contribution analysis with process tracing (CAPT) will serve as an overarching strategy for identifying, bounding, and contextualizing the contributions of TAG activities in shifting the Threshold Program's key outcomes of interest, which are defined in Figure II. The mixed-methods CAPT approach we propose draws on four research methods, two qualitative and two quantitative (Figure III.1). This research approach is particularly appropriate in this empirical setting, which lacks a comparison group (because all Kosovans are potential beneficiaries of TAG programming). The lack of a comparison rules out the use of experimental or quasi-experimental techniques for estimating causal impacts.[6] Contribution analysis can overcome this challenge by drawing on mixed research methods to assess the contribution of an intervention (even a society-wide one) to an observed outcome considering all other relevant factors or assumptions that may also have influenced the outcome.

The goal of a contribution analysis, like that of an experimental or quasi-experimental research design, is a substantiated statement on whether the studied intervention was successful and if so, to what degree. To conduct contribution analysis for TAG activities, we will dissect the existing project-wide and activity-specific TOCs node by node, identifying (1) other non-TAG activities or conditions that may have contributed to observed outcomes, (2) additional actors and assumptions that are not listed in the TOC that should be explicit to improve the TOC's accuracy, (3) whether desired TAG outcomes were reached, and (4) the specific contribution the intervention is claimed to produce for each outcome (Koleros and Mayne 2019). To facilitate the last step, we will use mixed-methods research tools and information on the temporal relationships between TAG activities and observed outcomes to assess each linkage and result in the TOC. For example, we will pair what we know about the TAG implementation timeline with quantitative data on citizens' open data engagement to examine whether the project's activities could

[6] The limited number of air quality sensors installed under EDC means that people living near a sensor benefit from the real-time air quality readings as well as the forecasts, whereas people with no sensor nearby only benefit from forecasts produced for all of Kosovo. Since the presence of an air quality sensor is necessary to assess predictive accuracy, we cannot appraise accuracy in areas not served by a sensor without using external data sources.

reasonably have driven observed changes in behaviors at a certain point in time. Such a temporal analysis can act as a falsification test to assess the project's contribution. If open data engagement began accelerating even before TAG implementation started, that would weaken the case for TAG having been the primary factor influencing data usage patterns.

Figure III.1. Organization of research methods supporting contribution analysis



Source: Mathematica evaluation team

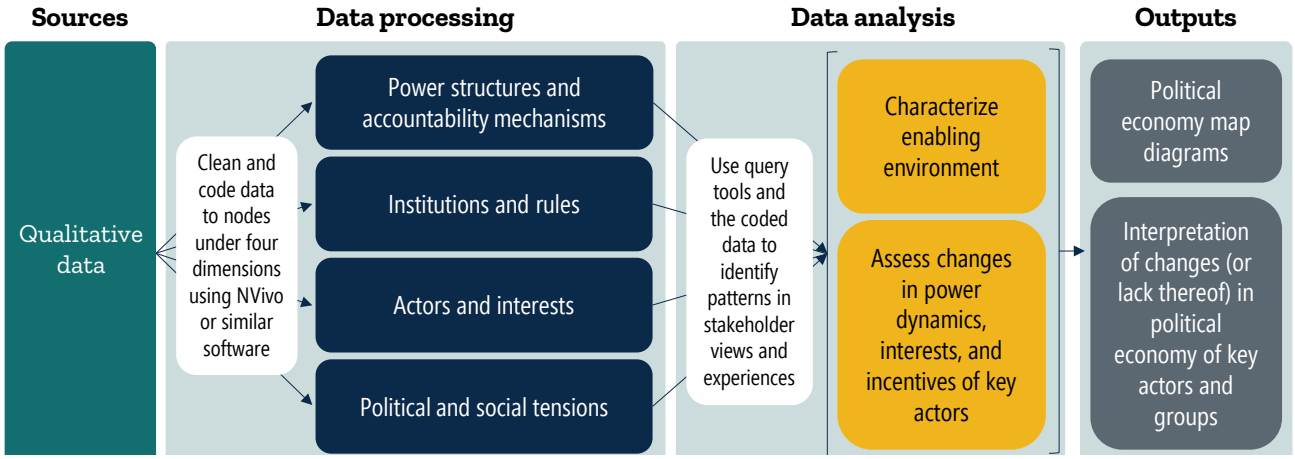
As part of our contribution analysis, we will draw on qualitative information to conduct **process tracing** on the TOC to examine each assumed causal link—for example, the claim that once the government of Kosovo, civil society, and the private sector can analyze data, those data will be communicated publicly. We will then examine and diagram the contributions of the intervention in each link in the pathway and assess whether outcomes can be sustained even after the intervention ends, considering all factors that influence sustainability. A key objective of CAPT is to provide a systematic approach to support causal statements in settings where there are several confounding factors. This is a key strength for evaluating TAG, given the number of other donors and initiatives that have concurrently operated in the judicial, environmental, and data liberalization spaces. For example, the EU has supported activities related to pollution abatement at Kosovo's largest thermal power plant, Kosovo B, which are likely to affect citizens' awareness of air quality and their level of interest in data on pollution. Using the CAPT approach to organize our research methods will enable us to isolate TAG's specific contribution to changing outcomes and impacts by systematically disentangling the effects of its activities from the effects of other related efforts.

As a pillar of the contribution analysis and process tracing approach, we will apply **political economy analysis** in the evaluations of PAJI and KODC to characterize the roles, relationships, incentives, and preferences of the constellation of actors involved in these activities. Using qualitative data, political economy analysis can illuminate the reasons why an intervention's desired outcomes—such as greater government transparency and public trust—were or were not fully achieved because of the distribution of power and interests across actors in the intervention's political space. A political economy analysis allows researchers and practitioners to draw valuable lessons from qualitative data across a variety of areas related to government and civil society. For example, researchers have used this type of analysis to understand health system resilience in Cameroon, Nepal, and South Africa during the COVID-19 pandemic (Williams 2022). Their analysis identified the factors in the enabling environment and in the interests of key actors, such as committed and well-informed political leadership, that supported the resilience of health systems. Pherali and Safar (2018) applied political economy analysis to understand why the education system of Afghanistan was providing space for radicalization of youth. The researchers found that the intersection of growing fundamentalism and state fragility (represented by deteriorating

security conditions, weak governance, and widespread corruption) provided an enabling environment for extremist actors to use educational spaces for radicalization.

In political economy analysis, researchers organize qualitative data into three key dimensions: power structures and accountability mechanisms, institutions and rules, and actors and interests (Warrener 2004). In addition to examining each dimension in this framework, we will include a fourth dimension: political and social tensions. We have successfully used this expanded political economy analysis approach in previous work with MCC, including on the El Salvador Investment Climate Project.[7] With information thus organized in a qualitative data processing and analysis software, such as NVivo, researchers can examine power dynamics and draw out the implications of those relationships (Figure III.2).

Figure III.2. Political economy analysis approach



Source: Mathematica evaluation team

Mapping political economy analysis allows us to document how different institutions and agents influence decision making and bargaining processes; examine the processes of institutional and policy change; and provide insights into what, how, and why change occurs in each sector. Using the Warrener (2004) and USAID (2018) frameworks, we will characterize the enabling environment of each intervention, assess the power dynamics, interests, and incentives of key players, and identify changes in those factors as they relate to the PAJI and KODC interventions. Data for the political economy analysis will be drawn from KIIs and focus group discussions (FGDs) with government employees and relevant CSOs. Key outputs of the political economy analysis will be political economy diagrams for each activity and interpretation of changes in the enabling environment, power dynamics, interests, and incentives of key players. Political economy analysis can help us interpret insights derived from quantitative methods, such as the reasons why citizens and CSOs may not download and use government data for analyses.

For regularly collected time-series quantitative data, we will apply **descriptive trends analysis** to understand the potential effect that TAG investments had on the trajectories of specific indicators or

[7] USAID recommends a similar approach, in which researchers organize information into foundational factors (deeply embedded structures in which agents interact), rules of the game (formal and informal institutions shaping incentives), and the here and now (immediate stimuli that affect behaviors and goals) (USAID 2018). We have found the Warrener approach works equally well for political economy analysis.

metrics. For example, we will compare values before and after TAG activities to determine if a shift in mean values or a change in slope might be attributable to the investments. Findings from descriptive analyses of quantitative data, in terms of timing and/or magnitude, will support the contribution analysis. For example, if we observe trend breaks in an indicator's time series that precede the start of an intervention, then we know TAG cannot be responsible. We would then take that information to enrich our qualitative data collection instruments to understand the cause of changes and aid in interpreting magnitudes. For example, consider a dataset of the daily number of downloads for an app that was supported by TAG. If our descriptive trends analysis reveals this number is large, then we would tailor our qualitative instruments to understand the source of this success and whether it might be transferable to other settings. Conversely, if there are few app downloads, then we would want to understand the impediments to broader citizen interest, and whether they reveal key learning opportunities for future program implementation. We will apply this research method to datasets such as website traffic, app downloads, and air quality sensor values, each described in greater detail in Section D. Under CAPT, descriptive trends analysis will help us validate perceptions and claims synthesized through our qualitative methods.

We will apply **qualitative thematic analysis and triangulation** across both qualitative and quantitative data sources. This approach involves thematically coding transcripts of KIIs, FGDs, and program documentation. We will triangulate evidence across sources to help interpret competing perspectives among stakeholders or information sources. Mapping actors involved in each activity and understanding their roles and responsibilities will enable us to assign information sources to different levels of importance and credibility. For example, implementers and government officials are likely to possess more credible knowledge on the technical factors affecting open data IT system maintenance than CSO stakeholders who do not work in cloud computing or with mainframe servers. Therefore, the perspective of implementers and government officials will carry more weight when we encounter contradictory explanations across information sources. As part of this process of analysis and triangulation, we will leverage available quantitative data sources to obtain stakeholders' independent appraisals of observed trends to inform our narrative explanation. As an example, we would examine trends in public sentiment of government effectiveness available through the Kosovo Public Pulse surveys and seek civil servant interviewees for their interpretation and explanation of any observed trends.

Lastly, we will use **correlation analysis** as part of our EDC evaluation to measure the correlation between air quality forecasts and actual air quality readings. This approach will enable us to test forecast accuracy, since forecasts are only useful if they accurately predict future air quality conditions.[8] This analysis will continue similar work done by NIRAS and Atmoterm (2021) which, because of the timing of our data collection, will enable us to compare whether the correlation between forecasts and measurements has changed after monitoring system ownership and responsibilities shifted from NIRAS to KHMI near the end of the Threshold Program.[9] We will report the correlation results using the Pearson correlation coefficient from analyzing both hourly and daily mean values for the four air quality parameters available both in real time and in forecasts.[10] Although forecast data are produced for the

[8] Given the counts of KHMI website visits and air quality app downloads, we have no reason to believe that forecasts would alter behavior on a large enough scale to impact actual emissions or fossil fuel combustion behavior and in turn affect monitor readings.

[9] This analysis may be temporally limited if KHMI does not continue generating AQ forecasts for the entirety of the data collection window.

[10] The four indicators are PM_{10}, $PM_{2.5}$, NO_2, and O_3.

entirety of Kosovo, we are geographically limited to areas immediately surrounding monitoring stations when conducting the correlation analysis, as only those sites have both forecast and actual air quality readings available. As with descriptive trends analysis, correlation analysis will illuminate findings delivered through qualitative analyses by quantifying relationships that interviewees and focus group discussants reference as reasons for or against behavior change.

D. Data sources and data collection

We will use a combination of primary and secondary data sources to answer the research questions. In this section, we describe the proposed quantitative and qualitative, data sources, along with their important characteristics (such as temporal coverage, granularity, available indicators, and so on) and any relevant limitations.

1. Quantitative data sources

A key focus of the evaluation will be understanding whether and how open data is used by civil society and relevant branches of the GoK. We propose using multiple quantitative data sources (Table III.8) to answer those questions. Where possible, we will disaggregate data by gender and ethnic group identification, but we note that many of the available quantitative data sources (for example, website visits or downloads of an open data set) are deidentified and therefore will not support GSI disaggregation. To the degree possible, the evaluation team will assess whether datasets allow for disaggregation by gender or other variables whether corresponding analysis and reports provide insights on issues related to gender or ethnic groups.

Table III.8. Quantitative data sources used by evaluation

Data source	Evaluation			
	PAJI	EDC	KODC	Cross-cutting
Google analytics (websites, apps)	✓	✓	✓	
Open datasets	✓	✓	✓	
Google Trends results	✓	✓	✓	✓
Air quality data (real-time and forecasts)		✓		
Judicial portal data (CTM and ODP)	✓			
Kosovo Public Pulse Surveys	✓		✓	✓
Social media content	✓	✓	✓	✓

The primary means of assessing interest in newly available open datasets and data portals is through the **Google Analytics data** generated for a website or a phone app (Android and iOS). Relevant indicators include the number of daily hits on a web page, the number of app downloads and regularity of app use, the number of daily downloads of a report or dataset, and the average number of (active) minutes spent browsing a web page. We anticipate obtaining Google Analytics data for select government websites (such as the CTM and ODP sites, the NIPH microsite, ERO's energy statistics site, and KHMI's air quality monitoring site), as well as for apps and websites developed through the four DigData challenges. For sites that do not participate in Google Analytics, we will work with whatever alternative data on usage statistics is available.

We will work closely with government stakeholders to monitor any new **open datasets that have been published on government servers** and to collect, where possible, download statistics for those data. For example, as part of their participation with the Threshold Program, KJC has published data that are now hosted on the KPC website. We may also work with KHMI, KEPA, NIPH, and other agencies to identify new open datasets for which we could collect download statistics. These data provide insight into the relative popularity of Threshold Program-supported data relative to other accessible datasets and will improve the search results of queries we perform for obtaining briefs, reports, studies, and any other documents that are predicated on TAG-supported data. As with the Google Analytics data, because users can download these datasets without creating a user profile, we will not be able to disaggregate usage statistics by gender or ethnicity.

We will use **Google Trends data** to assess changes in internet search intensity for keywords that are relevant to TAG programming, and to help diagnose whether such changes coincide with specific TAG-related dates, such as the launch of an information campaign or the release of a new data product. Searches will be conducted using keywords in both English and Albanian, and for Kosovo and neighboring countries like Albania, Montenegro, and Serbia (in local languages) to decipher if observed trends in Kosovo are country- or region-specific.[11,12] Google Trends data do not provide actual counts of search queries, but rather convert those counts into a 0 to 100 integer range by normalizing each week's values against the maximum weekly value in the defined period of interest and multiplying by 100. Consequently, there will always be at least one week with a Google Trends value of 100. We will use Google Trends data from January 2016 onward, which will provide us with several years of pre-intervention "baseline" data against which we can compare search habits during and after TAG.[13] Search intensities for all other weeks can then be interpreted as a percentage of that week of maximum search intensity. For example, search intensity for '*ndotja e ajrit*' ('air pollution' in Albanian) peaked in the week of January 28, 2018, as seen in Figure III.3. The week of April 19, 2020, saw the second largest number of search requests for this keyword, but was still only 49 percent of the search volume of the January 2018 peak. Data from multiple keywords cannot be combined because the values for each keyword are normalized against a denominator that Google does not share, so our analysis will look for patterns across related search terms to determine if there are consistent trends or an indication that preferred search terms have shifted over time (i.e., a gradual reduction in intensity of one keyword coincides with a gradual increase in a related term).[14]

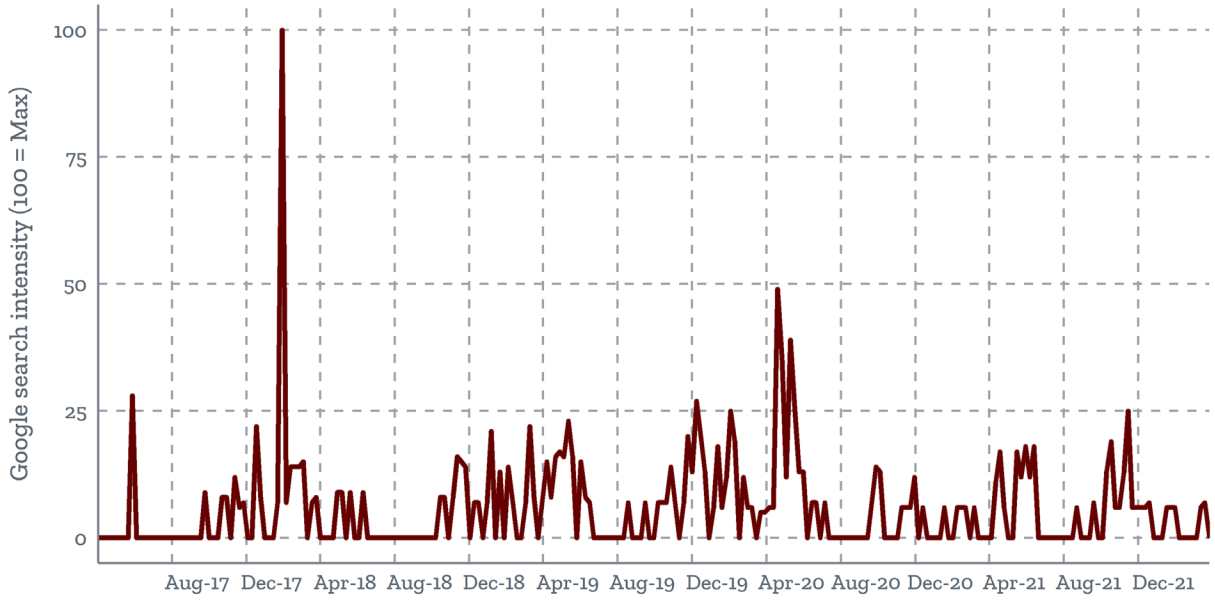
[11] The Google Trends site lists which cities have the highest interest levels for a specified search term, but not enough data are available to present time-series values for specific cities or regions within Kosovo. Google Trends results will consequently be presented for all of Kosovo. Mavragani and Ochoa (2019) offer several best practices in engaging Google Trends for public health surveillance purposes, such as joint queries that include common typos, how best to handle accent marks, and time period selection.

[12] Internet users who browse through a virtual private network (VPN) may adjust their country setting to appear to be based outside Kosovo. Those users may not appear in Google Trends data collected for Kosovo. For this to be a material risk to the evaluation, VPN users would have to systematically differ from non-VPN users in the timing and frequency of search requests related to air pollution, which we do not believe to be the case.

[13] Although Google Trends data are available from 2013 onward, Google updated their data collection system on January 1, 2016. By dropping pre-2016 values from our analysis, we will avoid concerns about methodological differences in data collection potentially contributing to any of our findings.

[14] We note that Google Trends data only capture Google searches, and therefore do not offer any insights into searches using alternative tools like Bing or DuckDuckGo. We do not have reason to believe that search platform preferences for searches related to air quality would have meaningfully changed during the evaluation period.

Figure III.3. Google Trends weekly search intensity for '*ndotja e ajrit*' ("air pollution" in Albanian) in Kosovo

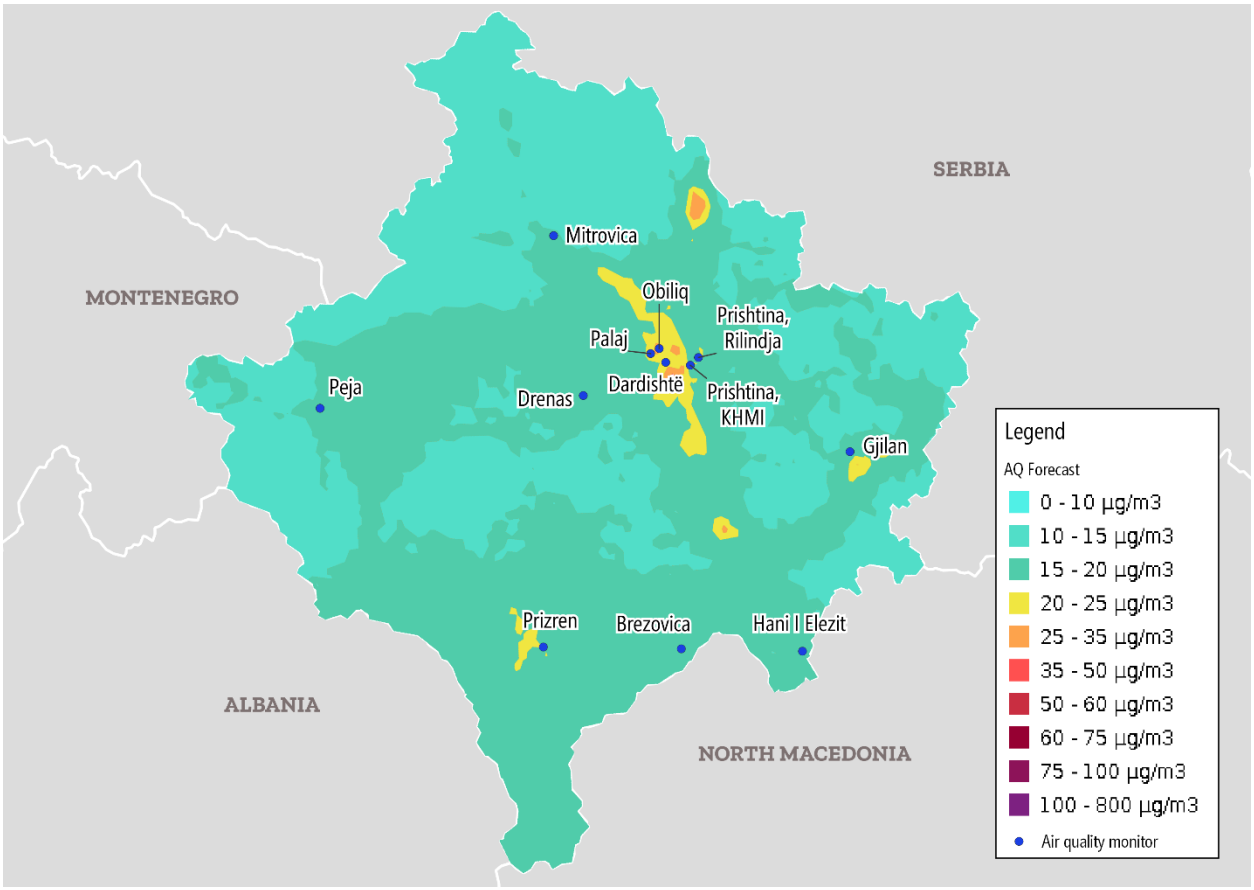


Source: Mathematica calculations using data from Google trends.

A key component of the EDC was the deployment of a network of air quality monitors at 13 sites (12 fixed locations and one mobile, displayed in Figure III.6) throughout the country.[15] We will use the **air quality data** recorded by the monitors to track changes in air quality and identify periods of pronounced air pollution. These data will also enable us to assess data missingness, which could be a relevant factor affecting trust levels and public confidence in the data. If missing values are more likely to coincide with times of the year, or parts of the day, when actual pollution levels tend to be highest, then such missingness could be just cause for citizens and organizations to be wary of the data's quality. Each monitor records hourly average values for six air quality parameters (particulate matter less than 10 microns in diameter (PM_{10}), particulate matter less than 2.5 microns in diameter ($PM_{2.5}$), nitrogen dioxide (NO_2), ozone (O_3), sulfur dioxide (SO_2), and carbon monoxide (CO)), and historical data are retrievable from a KHMI portal. EDC's TOC did not envision impacts on ambient air quality, and so our analysis of the data will not treat air pollution levels as an ultimate outcome of the project, but rather another input into the design of our qualitative instruments. For example, by scheduling FGDs around the time when indoor heating demand tends to be high, we may be able to identify extreme pollution episodes as reference points against which respondents can recount any defensive actions they undertook to reduce their pollution exposure. Related to actual air quality data, we will also use **air quality forecasts** that are produced for all of Kosovo and available for the air quality index, PM_{10}, $PM_{2.5}$, NO_2, and O_3. Forecasts are continuously updated and available with up to a three-day lead time. Figure III.4 is an illustrative forecast for $PM_{2.5}$ with reddening categories denoting successively more polluted areas according to particulate matter concentrations, measured in micrograms per cubic meter (μ/m^3).

[15] As of July 2022, the 12 stationary sites submitting data to KHMI's air quality network are Brezovica, Dardhishtë, Drenas, Gjilan, Hani I Elezi, Mitrovica, Obiliq, Palaj, Peja, Prishtina (KHMI), Prishtina (Rilindja), and Prizren. There is also one mobile site, which is primarily based within Prishtina.

Figure III.4. Illustrative map of $PM_{2.5}$ forecasts for Kosovo and locations of 12 air quality monitoring stations



Source: Mathematica illustration using $PM_{2.5}$ data downloaded from KHMI's web map service (WMS).

For our analysis of PAJI activities, we plan to use **aggregate data contained in the ODP, CTM, and CMIS**, as well as any available Google Analytics usage statistics for those services. Since these platforms have not yet been launched, their exact contents are not yet known nor are the specific variables that will be available for analysis.

The United Nations Development Programme (UNDP) has been conducting **Kosovo Public Pulse (KPP) Surveys** biannually since 2010. These opinion surveys are conducted with a representative sample of Kosovan adults and include questions related to satisfaction with leaders and institutions, perceptions of safety, and perceived rankings of the most important domestic problems. Since most of the survey questions appear in each survey round, the Public Pulse Surveys are the only recurring survey conducted in Kosovo to offer a longitudinal view of public attitudes on social matters. Although the Public Pulse survey questions do not perfectly align with the research objectives of the TAG evaluation (IDRA

2022),[16] we believe that monitoring trends will enable us to contextualize our data collection against broader developments in the citizenry, as well as stimulate conversations with stakeholders.[17]

We will also use **social media content, especially Facebook**, to track engagement with reports, analyses, and announcements related to TAG activities on the relevant platforms. Since Facebook's legal restrictions restrict web scraping, we will manually review pages and posts of organizations that have either produced TAG-related content or reshared others' content.

2. Qualitative data sources

Our qualitative data collection will be a combination of document review and in-person discussions with TAG stakeholders. In this section, we describe the qualitative data we will collect; Section E provides the specifics of documents or stakeholder group interviewees to where we define the activity-level evaluation designs.

We will use KODC **grantee documentation** to gather information on the implementation and impacts of the four DigData Challenges. Available documents include concept notes, milestone reports, annual reports, and budgets and workplans. Because these materials contain limited quantitative information (and the available data are not comparable across grantees or have distinct challenges), we will examine the qualitative information in the documents related to risks and assumptions, implementation challenges and successes, and achievements and plans for sustained impact. These documents will be part of our broader collection of **program documentation,** which will be crucial for the implementation analysis. Our efforts will focus on project timelines and workplans, progress reports, and analyses that were undertaken to guide key implementation decisions.

We will conduct **key informant interviews** with a range of stakeholders who, because of their involvement with the project, have distinct perspectives to share. In virtue of each activity contributing to the cross-cutting evaluation, KIIs will also support the cross-cutting evaluation. We will conduct KIIs with representatives from the public sector, civil society, academia, industry and professional groups, and activists. Most interviewees will be specifically relevant to one of the three Threshold Program activities, but a subset will be interviewed for their experience relating to two or more of the activities. In the

[16] For example, one question asks respondents about their "awareness regarding their right to live in a healthy and clean environment" (UNDP Kosovo 2022), which does not directly respond to potentially relevant evaluation RQs 5, 6, or 9.

[17] We note three significant limitations with these data. First, the survey sample is a representative sample of the Kosovan population, not a stratified sample based on recent, direct experience with specific government services (IDRA 2022). Consider the judiciary as an example. Assuming that a minority of citizens are personally involved in a lawsuit at any given point in time, the KPP perceptions may be more a reflection of media portrayals of landmark cases since relatively few respondents are drawing upon direct experience when responding. As a result, aggregate numbers are more likely to represent attitudes about salient media events than firsthand (or related) experience as a plaintiff or defendant in an ongoing or recent legal case. A second limitation is the role that COVID-19 has played on many public perceptions, demonstrably affecting several indicators that would otherwise be of use to the evaluation. We have analyzed some of the most pertinent indicators and found large swings that coincide with the onset of COVID-19, which means that changes in public attitudes could never be linked exclusively, or primarily, to TAG activities. Despite these limitations, we believe the Public Pulse data can support the contribution analysis, especially as two more rounds of survey data may become available during the evaluation. A third limitation is that KPP survey questions often align poorly with the objectives of this evaluation. For example, the most pertinent survey questions for EDC included "C5. To what extent are you aware of potential environmental threats on your health/your family health?" and "C5a. How much do you think you know your rights to live in a healthy and cleaner environment?" (UNDP 2021).

following section, we provide more details about the stakeholders we intend to interview and describe how such conversations will help answer our research questions. Our KIIs will emphasize understanding the GSI implications of TAG activities, and as such we will interview CSOs whose mandate is specifically aligned with representing women and ethnic minority groups.

We also will use **focus group discussions** to catalyze dialogue among several multiple activity participants in the hopes of generating insights that may not have been possible through one-on-one conversations. Each group will consist of 6–10 participants and address relevant themes for the PAJI and EDC activities, which we respectively describe in Sections E.1 and E.2. For each FGD, at least two members of our research team will facilitate and guide the conversation. We will also examine the feasibility of conducting two cross-cutting FGDs with non-participant citizens, one in Pristina and one in another city or town, to assess the extent to which information from PAJI, EDC, and KODC activities is reaching people outside professional networks and how that information may shape their knowledge, attitudes, and practices.

We will rely on **reports and other publications** from government agencies, civil society, and academics. We will examine the relevance of data access to the insights they generated, and whether policy recommendations could have been generated absent access to those data. As with other secondary data sources, we will attempt to time analysis of program documents to precede interviews so that interview protocols can be informed by any findings coming from analyzing these documents.

Lastly, **Facebook and other social media** data can be used to track public engagement with CSOs' reports and analyses. We will also use social media data to assess how organizations and government agencies are or are not publicly communicating with one another as well as with society at large. As with a report or study, we can code social media posts and content in order to extract key themes that help contextualize how CSOs are using data, the challenges they have faced in communicating results, and their experience in coordinating with government counterparts to advance recommended social or policy reforms. Social media content can help us understand the depth with which organizations are involved in judicial or open data topics and inform the process of assigning importance and credibility to actors in cases of conflicting information, as described earlier in our qualitative thematic analysis and triangulation.

E. Activity-level evaluation designs

In this section, we offer a detailed view of our proposed evaluation designs for the three activities and the cross-cutting program-wide evaluation. For each activity and the cross-cutting evaluation, we describe study samples, data sources and collection strategies, timelines and exposure periods, analysis plans and challenges and limitations.

1. PAJI evaluation

A key challenge for the PAJI evaluation is the delayed activity implementation, which may compress the activity's timeline and reduce the scope of the activity and will limit our ability to observe longer-term outcomes within the evaluation period. We will take advantage of the available time and data sources to build an evaluation of PAJI that provides the clearest view of the activity's implementation and contribution to targeted outcomes. Table III.9 shows the proposed method, key indicators, and data sources by research question (RQ) for the PAJI evaluation.

Table III.9. Overview of PAJI evaluation design

Evaluation methods	Key indicators	Data sources
RQ1. Was the activity implemented according to plan (in terms of quantity and quality of outputs)?		
• Implementation analysis • Qualitative thematic analysis and triangulation	• Key indicators listed in the M&E plan, such as judicial actors trained on data communications (PA 5.0) • Stakeholders' reported use of the judicial portal and the data published there	• Implementation documentation (including implementation reports, PAJI ToR, work plans, program M&E Framework, internal reports generated by PAJI management, MFK M&E data in ITT) • KIIs and FGDs
RQ2. Did the activity achieve its targeted outcomes, particularly its stated objective, in the timeframe and magnitude expected? Why or why not?		
• Implementation analysis • Descriptive trends analysis • Qualitative thematic analysis and triangulation	• Key program outcomes listed in the M&E plan achieved, such as reduced judicial processing time (PA 7.0) • Improved access to justice for citizens • Data used for decision making by government and other relevant actors • Increased advocacy, collaboration and communication between judiciary and civil society • Changes in judicial efficiency	• Implementation documentation • Google Trends data • Data from judicial portals created by PAJI activity • MCC Kosovo Scorecards[1] and ITT • KIIs and FGDs
RQ3a. Is there any increase in the Government's use of analyses done by non-government entities, both analyses supported by MFK and those in general? Is there any evidence these types of analyses can factor into any policy decision-making? **RQ3b. Did publishing judicial data through relevant government websites result in increased analysis done by NGOs?** **RQ3c. Did PAJI result in increased engagement between government and civil society/media?**		
• Descriptive trends analysis • Political economy analysis • Qualitative thematic analysis and triangulation	• Government agencies' use of NGO analysis • Government agencies' use of judicial data to inform decision-making • NGOs' access and analysis of newly available data on government websites and portals • Improved relationship between GOK and civil society due to increased data access and government transparency • Government employees' perception of data as public good or resource to be shared	• Implementation documentation (see above) • Data from online portals created by PAJI activity (usage statistics—breakdown by mobile, by desktop computer, timing of use) • Report downloads from ODP and CTM (PDF and Excel) • Google Trends data • KIIs and FGDs
RQ4. How has the project contributed to citizens' use of judicial data in advocating for change?		
• Descriptive trends analysis • Political economy analysis • Qualitative thematic analysis and triangulation	• Perception of citizens' use of data to advocate for change • CSOs and citizens' use of data to advocate for change • Changes in organization of power, decision making, and economic resources among key actors (including changes in relative power of CSOs and citizens to advocate policy change) • CSOs' use of disaggregated data to identify and remedy cross-group differences in access to justice	• Google Trends data • KIIs and FGDs • Social media content

Evaluation methods	Key indicators	Data sources
RQ5. Did PAJI contribute to increasing trust and understanding of the judiciary system's functions?		
• Political economy analysis • Qualitative thematic analysis and triangulation	• Public perceptions of trust in government with respect to judicial disclosures and policymaking • Changes in organization of power and decision making among key actors	• UNDP Public Pulse Survey data • KIIs and FGDs

[1] Since MCC Kosovo Scorecards were maintained only until to 2019, we will use the scorecards' sources for 2020, 2021, and 2022: IMF WEO (for Inflation, Fiscal Policy), Freedom House/CLD (Political Rights, Civil Liberties, Freedom of Information), World Bank/Brookings WGI (Control of Corruption, Regulatory Quality, Government Effectiveness, Rule of Law, Gender in the Economy), IFAD/IFC (Land Rights and Access, Access to Credit, Business Start-Up), Heritage Foundation (Trade Policy), WHO/UNICEF (Health Expenditures, Immunization Rates), UNESCO (Primary Education Expenditures, Girls' Secondary Education Enrollment Rate), CIESIN/YCELP (Natural Resource Protection, Child Health).

a. Methodology

Our PAJI evaluation will draw on each of the research methods listed in the first column of Table III.9. Across all research questions, we will process data from KIIs and FGDs with qualitative thematic analysis and triangulation to assess the credibility of assertions expressed in the data and the differences and alignments in respondents' viewpoints and experiences.

We will apply an implementation analysis framework to answer RQ1 and RQ2 and uncover factors that contributed to any deviations in actual implementation against project plans. We will draw upon the project documentation that is available, including implementers' progress reports, implementers' annual reports, and minutes from donor coordination meetings as a starting point for follow-up discussions with MFK staff and implementers. For RQ2, which asks whether PAJI achieved its targeted outcomes, we will collect qualitative data from key informants and focus group discussants, including government stakeholders and civil society actors who are active in the rule of law space. To support our descriptive trends analysis for RQ2, we will also draw on Google Trends data, CTM and ODP data, and data from the MCC scorecards and their sources.

RQ3 examines the degree to which the provision of judicial data has altered the ability of CSOs and the government to conduct new analyses based on those data, whether such analyses are informing policymaking, and whether PAJI resulted in increased engagement between government and civil society or the media. Although we will report the available quantitative indicators on the number of analyses based on the ODP data, and trends in web traffic for the ODP and CTM portals, we will primarily answer this RQ by interviewing government officials and representatives from rule of law CSOs and media outlets. We will apply political economy analysis to interpret changes in stakeholders' power, institutional arrangements, and interests, and will use qualitative thematic analysis and triangulation to help determine the strength of different viewpoints expressed in the qualitative data. Combining descriptive trends analysis with qualitative thematic analysis and triangulation and political economy analysis will help us trace the links between results stages and identify the contribution of PAJI in light of concurrent initiatives to improve judicial transparency conditions more broadly.

RQ4 is closely related to RQ3 and examines how the PAJI activity has contributed to citizens using data to advocate for change. We will answer this question through interviews with CSOs and rule of law and transparency activists, FGDs with users of the judicial data, and reviewing social media posts. We will rely on political economy analysis, qualitative thematic analysis and triangulation, and descriptive trends analysis to process and interpret data collected to answer this question. Key indicators for answering this

question include documented use of judicial data to advocate for judicial process improvements and barriers that are identified as major obstacles to effective citizen advocacy. As with RQ3, political economy analysis will allow us to catalog where the different stakeholder groups and institutions stand in terms of their level of support for PAJI-supported tools and their level of influence, and descriptive trends analysis will help us identify changes in citizen engagement and advocacy. Combining the two analyses with qualitative thematic analysis and triangulation will be instrumental in interpreting the causes for any changes we find, including citizen engagement and trust outcomes associated with PAJI.

Lastly, in RQ5 we will assess whether PAJI improved the public's trust and understanding of the Kosovan judiciary. We will consult private citizens, CSO staff, and government officials to learn if they believe that public trust in the GoK has changed as a result of PAJI activities. As with previous questions, we will process qualitative data with political economy analysis and with qualitative thematic analysis and triangulation. Political economy analysis will help us assess the degree to which increased transparency contributes to changes public trust and understanding—and what barriers and challenges to change in those outcomes might remain. The method will also help us understand how increased judicial transparency may or may not contribute to efficiency of judges and the judicial system, highlighting how the incentives or power dynamics of key stakeholders may influence that outcome.

b. Study sample

Since the ODP developed under the PAJI activity was designed to be accessible to all Kosovans, the sample frame technically comprises all Kosovan citizens. The CTM sample frame is parties to cases being processed in the Kosovar judiciary, and the PAJI outreach and communication strategies are generally restricted to actors in the judiciary or in NGOs. For the evaluation, we will primarily focus qualitative data collection on individuals and organizations who have directly engaged with tools supported through the Threshold Program. We will define this group by identifying organizations and individuals who have been trained on (or developed or manage) the data platforms or who have used them as private citizens or civil society actors.

c. Primary data collection

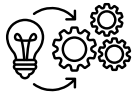
We will collect primary data for the PAJI evaluation through KIIs and FGDs. In Table III.10, we list potential interviewees or focus group populations, the key themes that would be addressed in those meetings, and the number of KIIs or FGDs we plan to conduct with each type of stakeholder. To incorporate additional or emerging perspectives, in some cases we identify more potential participants than we expect to need. We also provide a target range of interviewees for some stakeholder groups, assuming that in some cases we may be able to engage two individuals in one institution in one interview—for example, two team members who report on judicial issues at a media outlet volunteer to be interviewed together. As the PAJI activity engages a wide variety of stakeholder institutions and groups, we will pursue a broad qualitative primary data collection strategy targeting between 19 and 29 interviewees and 4 focus groups comprising up to 27 individuals. One of the motivations in conducting multiple interviews within a given stakeholder group, as well as across stakeholder groups, is to corroborate across sources to assess where there is consensus in perceptions, experience, or outcomes. Corroboration will be particularly helpful when diagnosing which factors mediated PAJI's impact, and which factors were critical in determining why PAJI did or did not achieve its objectives. In addition to corroboration, we believe interviewing multiple interviewees in one stakeholder group can allow us to reach saturation—a point at which we have collected all major (even diverging) viewpoints and each additional interview yields diminishing marginal value in data. This may allow us to cancel interviews

with the fourth or fifth interviewees in a given stakeholder group when we feel confident that we have collected adequate data from other interviewees who had more proximity to—and information on—the activity than the next potential interviewee.

Since data available through the ODP is deidentified and cannot support any disaggregated analyses, we will obtain participants' perspectives on how PAJI may have had differential effects across gender and social groups through the KIIs and FGDs. To increase the diversity of perspectives available from participants, we will ensure our participant recruitment process results in speaking with individuals who are representative of the project's priority groups for inclusion.

We plan to recruit three to five judges and clerks from Pristina Basic Court, the Court of Appeal, and Peja Basic Court to gather insights on the ease of use of the CMIS and CTM platforms and the perceived relationship (and level of trust) between citizens and the judiciary. To recruit these members of the judiciary, we will work with our PAJI evaluation lead, Dr. Ariana Qosaj-Mustafa, to identify judges in these courts through the ODP, and will then ask willing judges to suggest one to two clerks with whom we can request an interview as well. These clerks could serve the same judges we will interview or may serve other judges, but our priority will be identifying clerks with longer tenure to understand their experiences across multiple data management systems.

Table III.10. List of potential interviewees and FGD participants for PAJI evaluation

Stakeholder type	Potential participants	Illustrative themes addressed	Target number
Key informant interviews			
MCC/MFK 	• MCC* • MFK*	• Project implementation and fidelity to initial plans • Perceived collaboration between government and civil society/media • Interpretation of trends observed in quantitative data sources	2–4 interviewees
GoK agencies 	• Kosovo Judicial Council* • Kosovo Prosecutorial Council* • Ministry of Justice • Agency of Gender Equality* • Ministry of Internal Affairs • Kosovo Ombudsperson Institution	• Ease of use and transfer of the ODP and CTM to relevant government offices • Perceived relationship and level of engagement between government and civil society • Use of open data and relevant analyses in decision making • Inter-agency communication and collaboration on judicial issues	3–5
International organizations 	• Council of Europe European Commission for the Efficiency of Justice (CEPEJ)	• Information on past and ongoing projects with the judiciary • Perception of public trust in judiciary	1
Justice system 	• Clerks and judges from: • Pristina Basic Court • Court of Appeal • Peja Basic Court	• Ease of use of the CMIS and CTM • Perceived relationship (and level of trust) between citizens and the judiciary	3–5
Other donors 	• UNDP • USAID Justice Program in Kosovo • Norwegian Embassy	• Connections between PAJI and other donor-funded projects • Perceptions of PAJI effectiveness and key results • Contribution of donor coordination on PAJI impacts	3
Implementers 	• Kosovo Legal Services Company • B&S Europe • InfoSoft Systems	• Implementation challenges and threats to project sustainability • Changes to PAJI programming and anticipated effects • Interpretation of trends observed in quantitative data sources	3-5

Stakeholder type	Potential participants	Illustrative themes addressed	Target number
Media 	• Radio TV of Kosovo* • Radio TV 21* • Balkan Investigative Reporting Network • KTV* • Drejtesia ne Kosove • Betimi per Drejtesi	• Perceptions of availability of judicial data for analysis and reporting • Use of proprietary analysis (or analysis conducted by CSOs) of judicial data • Perception of citizen engagement with judiciary and judicial issues	3-5
Academics/ Researchers 	• Two Kosovo-based academics with expertise in the judiciary and rule of law issues (names removed for this report)	• Ease of citizen and media engagement with the judiciary • Additional needs for using judicial data to affect social change • Perceptions of public trust in and understanding of GOK's function	2
Focus groups			
Mixed group of 2 prosecutors each from Pristina and Gjakova offices with different levels of seniority		• Ease of use of CTM platform • Perceived value of open judicial data • Remaining barriers to judicial transparency	4 participants
Mixed groups consisting of CSOs/NGOs working in the judicial area and related social advocacy areas, including Kosovo Law Institute, Legal Aid Agency, Group for Legal and Political Studies, Open Data Kosovo, Democracy Plus (D+), Center for Social Group Development, FOL, Çohu, Kosovo Women's Network, Center for Gender Studies, Roma, Ashkali, and Egyptian Women Network, Youth Initiative for Human Rights, Aktiv, Kosovo Women's Chamber of Commerce – G7, including organizations that were involved in the design and implementation of PAJI and those that were not		• Use of judicial data in social advocacy • Availability and accessibility of judicial information (for women and ethnic minority populations in particular); perceived transparency of judiciary among those groups and in general • Perceived relationship and level of engagement between government and civil society • Interpretation of trends observed in quantitative data sources	2 focus groups, with 6–8 participants each
Mixed group of Kosovan citizens drawing from commenters on articles that use ODP data and from individuals with active cases who interact with KJC and KPC online		• Usefulness and useability of available judicial data platforms • Actions undertaken with newly available judicial information • Trust in government, the judiciary	5–7 participants

Note: CSO = Civil society organization; MFK = Millennium Foundation Kosovo; NGO = Non-governmental organization. Asterisks signal potential participants who will be interviewed once for their experiences and perceptions across multiple TAG activities.

FGDs will supplement KIIs in our primary data collection by allowing individuals with similar relationships to the PAJI to generate richer insights through discussion and interplay of experiences and ideas. We will conduct three types of FGDs to efficiently gather rich information from groups that share similar but somewhat distinct backgrounds and experiences with the judiciary and judicial data. First, we will work with KPC to hold a focus group with prosecutors in two cities and at two levels of seniority, to get a cross-section of prosecutors' experience with judicial data, transparency, and PAJI's activities. Second, we will gather insights on PAJI from two FGDs with representatives from two types of CSOs/NGOs: (1) organizations focused on judicial issues, and (2) social advocacy organizations with linkages to judicial transparency and citizen empowerment. Third, we will recruit citizens who have engaged in judicial issues through media or through government outreach to understand their access to judicial data and changes in their behaviors and perceptions.

d. Secondary data collection

The PAJI evaluation will draw upon many of the secondary data sources described in Chapter III, Section D.1. Table III.11 summarizes the secondary data sources we will rely on along with their purposes.

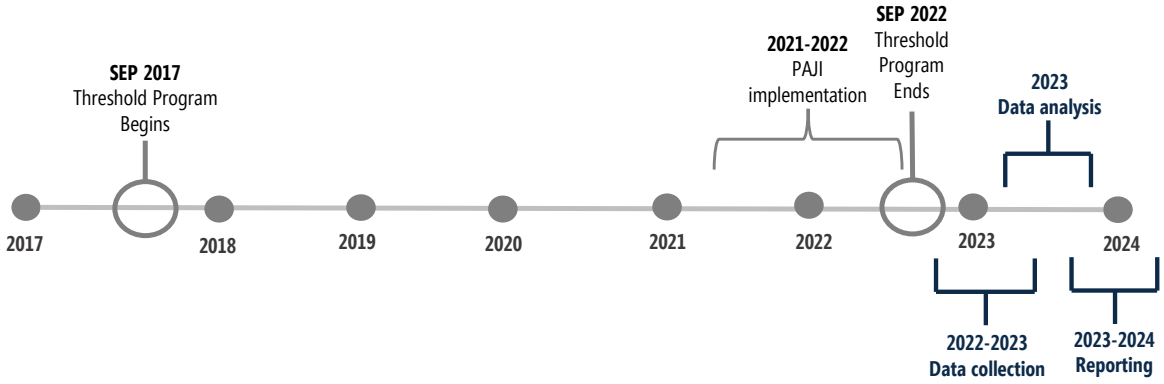
Table III.11. Overview of secondary data sources supporting the PAJI evaluation

Data source	Contents	Purpose
Implementation documentation	Reports, meeting minutes, project summary documents, terms of references, indicator tracking tables, Threshold Program close-out reports, and any other documents encompassing PAJI plans or actual implementation.	• Determining whether PAJI achieved key outputs and outcomes • Identifying differences between plans and actual implementation
Data from ODP, CTM, and CMIS	Download counts and daily traffic volumes based on Google Analytics data or other software analytics output.	• Tracking changes over time in public use of judicial data • Assessing when open data products were released been released and their popularity over time
Kosovo Public Pulse Surveys	Measurements of citizen trust in key judicial and political institutions over time	• Contextualizing qualitative data with longer-term trends in citizen perspectives
Google Trends data	Queries will be conducted in Albanian and English for pertinent search terms such as 'efficiency of courts' and 'judicial transparency'. The set of search terms will be finalized through consultation with MFK and CSOs.	• Detecting whether citizens' demand for judicial information has increased • Providing context for whether project achieved targeted outcomes
Facebook posts	Organizations/government agencies posting/reposting about judicial issues or data, public engagement levels with posts over time.	• Monitoring how organizations/individuals engage with judicial data and analyses • Discerning whether judicial data are supporting citizens' advocacy efforts
General usage statistics for KPC, KJC, and MoJ websites	Download counts and daily traffic volumes based on Google Analytics data or other software analytics output.	• Tracking changes over time in public interest in judicial issues • Contextualizing engagement between civil society and the GoK

e. Timeline and exposure period

Changes in the perceptions and behaviors of stakeholders involved in the PAJI activity may not begin to manifest until six months or more after the data systems are launched and operational. Because the implementation of PAJI was delayed (and will only end when the Threshold Program concludes), we will conduct most data collection work for PAJI in 2023 to increase the activity's exposure period, the fruitfulness of qualitative responses, and the number of observations in quantitative datasets. However, we will begin interviews with MFK staff and PAJI implementers as soon as possible in late 2022 to capture their implementation experiences while they can readily recall their work in the Threshold Program (see Figure III.5).

Figure III.5. PAJI data collection and analysis timeline



As indicated in the evaluability assessment (Chapter III, Section B), we expect that changes in participants' behaviors (such as CSOs using new analyses to engage with government on judicial issues) depend on changes in those participants' capabilities, opportunities, and motivations, which may take time to accrue. By postponing most PAJI data collection until Quarter 1 (Q1) of 2023, we can effectively increase the duration of exposure for most PAJI stakeholders and beneficiaries (government agencies, CSOs, the media, and citizens) to the activity's data products. We anticipate that this time period will allow stakeholders and participants to experience more changes in their capabilities, opportunities, motivations, and behaviors, so we can better assess the contribution of the activity. We will analyze the primary and secondary data in Q1/Q2 of 2023 and share our final results in reports and presentations between Q4 of 2023 and Q1 of 2024.

f. Analysis plan

For all KIIs and FGDs, we will develop guides and protocols in advance and submit them to MCC/MFK for review. The content of those guides and protocols will be informed by our review of the relevant literature on open judicial data systems and information communication, project documents, interviews conducted by Mathematica during a June 2022 mission trip, and analysis of any available quantitative datasets prior to the KII/FGD. For all interviews and focus groups, we will seek participants' consent to record the conversation. Individuals who are not comfortable speaking in English will be interviewed in Albanian, with interview transcripts or notes translated to English and individuals' statements anonymized. We will use respondent codes to ensure that any notable pull-out quotes or statements are linked to the correct respondent in line with best practices in reproducible research.

All qualitative data sources—KIIs, FGDs, program documentation, and reports and analyses by CSOs and government agencies—will be coded in NVivo or another qualitative software to systematically extract key themes that will facilitate comprehensively answering all PAJI-related research questions. Our processing of qualitative data will follow the qualitative thematic assessment and triangulation approach described earlier. Specifically, we will identify the incentives and levels of information access that each interviewee and or group of discussants has in order to assess the credibility of their assertions and triangulate their assertions, viewpoints, and experiences with those presented by other interviewees and discussants. We will use political economy analysis to understand inter-agency relationships and broader political forces operating outside of the agencies directly and indirectly involved with CTM and ODP.

We have earlier outlined the quantitative datasets for which we will apply descriptive trends analyses, which includes time-series data on Google Trends search intensity, and website traffic on portals related to PAJI. Among those datasets, only the Google Trends data are available prior to the start of PAJI activities. We will use the full duration of data and compare the timings of above-average search activity against important judicial events (such as KJC's announcement of the ODP launch or a prominent CSO's publication of an investigation into judicial backlog) to understand the degree to which contemporaneous media coverage and analyses drive internet search patterns for judiciary-related keywords. If we observe a lack of association between the two, suggesting that the visible actions of government, media, and civil society are not a primary determinant of aggregate search behavior, we will identify episodes of above-average search activity and ask KII and FGD participants for their perceptions of other factors driving internet searches of judicial information.

For time-series data available on ODP and CTM website traffic, we will conduct descriptive trends analysis. As only post-treatment data are available, we will not be able to compare recent slopes and means against pre-treatment values, but with the postponed PAJI data collection strategy described above, we will be able to document how usage patterns have changed in the six months after the portal launched, helping us interpret how PAJI tasks like outreach and training for civil society and the release of new videos on MFK's YouTube channel may affect CSO and citizen engagement.

Finally, we will use data from the UNDP Public Pulse Surveys related to citizen satisfaction in their government and the judiciary to contextualize trends in media coverage, trends in social media engagement and website traffic, and qualitative insights we gain on judicial transparency and data use and advocacy for judicial reform.

The CAPT framework described in Chapter III, Section C will be the overarching analytical approach used in answering PAJI evaluation research questions. A key advantage of CAPT is its role in integrating insights from other research methods. For example, applying a descriptive trends analysis on website traffic statistics will uncover usage trends (e.g., growth, shrinkage, or stagnation), the timing of peak interest, and the types of content garnering the most engagement from social media users. Such results can be directly interpreted through a CAPT lens or can be overlaid with other data sources such as interviews and FGDs to diagnose potential causes of such patterns and to highlight PAJI's contribution to any observed dynamics in collaboration, analysis, and discussion of judicial data.

g. Challenges and limitations

Because PAJI implementation was delayed, even an extended data collection period is unlikely to allow us to gather as much information on the activity's long-term effects (for example, in terms of behavior change) as if PAJI implementation had occurred earlier in the Threshold Program. While our evaluation

cannot predict future outcomes, we will seek to address the issue of a limited program exposure period by collecting qualitative information on stakeholders' perceptions of the current state of judicial transparency and on their expectations for changes in transparency, collaboration, and trust in the next three to five years.

2. EDC evaluation

In this section, we describe our proposed performance evaluation of the EDC (Table III.12). We outline the methods we will use to answer each research question, define the study sample, specify our primary and secondary data sources, and indicate the timeline for our EDC evaluation activities. We then explain our analysis approach for each RQ and list key challenges and limitations that are specific to this evaluation.

Table III.12. Overview of EDC evaluation design

Evaluation methodology	Key indicators	Proposed data sources
RQ1. Was the activity implemented according to plan (in terms of quantity and quality of outputs)?		
• Implementation analysis • Qualitative thematic analysis and triangulation	• Key indicators listed in the M&E plan, such as the number of partnerships between GoK and civil society/media (EDC 15.0) and the number of behavioral change activities conducted (EDC 2.0) • Stakeholders' reported use of data portals	• Implementation documentation (including implementation reports, terms of references, work plans, and project M&E framework) • KIIs
RQ2. Did the activity achieve its targeted outcomes, particularly its stated objective, in the timeframe and magnitude expected? Why or why not?		
• Implementation analysis • Descriptive trends analysis • Qualitative thematic analysis and triangulation	• Key program outcomes listed in the M&E plan achieved • Data used for decision making by citizens, health advisories, government, and other relevant actors • Increased advocacy, collaboration and communication between government actors and civil society	• Google Trends data • Data from online portals created by TAG Project (e.g., usage statistics by timing and device type, browsing duration) • MCC Kosovo Scorecards[1] and ITT • Implementation documentation • KIIs
RQ3a. Is there any increase in the Government's use of analyses done by non-government entities, both analyses supported by MFK and those in general? Is there any evidence these types of analyses can factor into any policy decision-making? **RQ3b. Did the dissemination of air quality data through government websites affect activities by NGOs/CSOs, and if so, then why?** **RQ3c. Did EDC result in increased engagement between government and civil society/media?**		
• Descriptive trends analysis • Qualitative thematic analysis and triangulation	• Government agencies' use of NGO analysis • Government agencies' use of data to inform decision-making • NGOs' access and analysis of air quality data and experiences in interpreting results • Any changes in relationship between GoK and civil society due to increased data access and government transparency • Government employees' perception of data as public good or resource to be shared	• Data from online portals created by TAG Project (usage statistics – breakdown by mobile, by desktop computer, timing of when they're looking – when pollution levels are particularly high) • Google Analytics and Google Trends data from websites and apps • Implementation documentation • KIIs

Evaluation methodology	Key indicators	Proposed data sources
RQ4. How has the project contributed to citizens' use of environmental data in advocating for change?		
• Descriptive trends analysis • Qualitative thematic analysis and triangulation	• Perceptions and experience of citizens' use of data to advocate for change • Barriers to citizen advocacy for improved environmental outcomes	• Data from online portals and phone apps created by TAG Project • NIPH portal data • Google Trends data • Facebook posts • KIIs • FGDs
RQ5. Does the existence of transparent, government produced air quality data, health advisories and a national outreach and behavior change campaign create enough incentive for civilians to change their behavior (e.g., take actions to reduce the negative health impacts of air pollution)? If evidence of changed behaviors exists, who is adapting, how have they adapted and why?		
• Descriptive trends analysis • Qualitative thematic analysis and triangulation	• Frequency and platform from which the air quality portals are accessed • Relationship between air quality and mobility patterns • Associations between media reports of air quality/pollution, use of the air quality portal data, and behavior change • Awareness of potential environmental health threats • Adoption of clean fuels and technologies for cooking, light, and heating	• KHMI AQ monitor readings and forecasts • Facebook posts • KIIs
RQ6a. Does the existence of transparent, government produced air quality data reduce the adversarial relationship between civil society and the GOK? If yes, whose attitudes and behaviors are likely contributors to these reductions? **RQ6b. Has inter-ministerial communication changed, e.g., between KEPA and NIPH, and if so, why and how?** **RQ6c. Is the air pollution data available on a continuously updated basis? How accurate are the air pollution forecasts provided through the NIPH portal? What percentage of time does air quality exceed given thresholds?**		
• Correlation analysis • Qualitative thematic analysis and triangulation	• Changes in the adversarial relationship between GOK and civil society • Frequency, type of communication among government agencies • Portal and/or station monitor downtime • Correlation coefficient estimates and root mean square error (RMSE) differences between actual and forecast pollution readings	• NIPH portal data and website performance statistics • KHMI AQ monitor readings and forecasts • Implementation documentation • KIIs
RQ7. Did EDC contribute to increased trust and understanding of government's function?		
• Qualitative thematic analysis and triangulation	• Public perceptions of trust in government with respect to environmental disclosure and environmental policymaking	• KIIs • FGDs

[1] As with PAJI, we will refer to the data sources used for MCC Kosovo Scorecards, which ended in 2019.

a. Methodology

Our EDC evaluation will draw on each of the research methods listed in the first column of Table III.11. Across all research questions, we will process data from KIIs and FGDs with qualitative thematic analysis and triangulation to assess the credibility of assertions expressed in the data and the differences and alignments in respondents' viewpoints and experiences.

We will apply an implementation analysis framework to answer RQ1 and RQ2 and uncover factors that contributed to any deviations in actual implementation against project plans. We will draw upon the project documentation that is available, including implementers' progress reports, implementers' annual reports, and minutes from donor coordination meetings as a starting point for follow-up discussions with MFK staff and implementers. For RQ2, which asks whether EDC achieved its targeted outcomes, our set of key informants will also include government stakeholders and civil society actors who are active in the environmental advocacy space. While quantitative indicator sourced from the ITT and statistics on data portal usage will help clarify whether EDC achieved its outcomes, interviewees' perspectives will be crucial in helping us understand why the outcomes were (or were not) achieved.

RQ3 examines the degree to which the provision of air quality data has altered the ability of CSOs and the government to conduct new analyses based on those data, and whether such analyses are informing policymaking. While we will report the available quantitative indicators on the number of analyses based on the air quality data, and trends in web traffic data for sites hosting air quality data, we will primarily answer this RQ by interviewing government officials and representatives from environmental CSOs. We will apply CAPT to interpret and bound the influence that EDC has had on these outcomes, given the presence of concurrent initiatives to address air quality and environmental conditions more broadly.

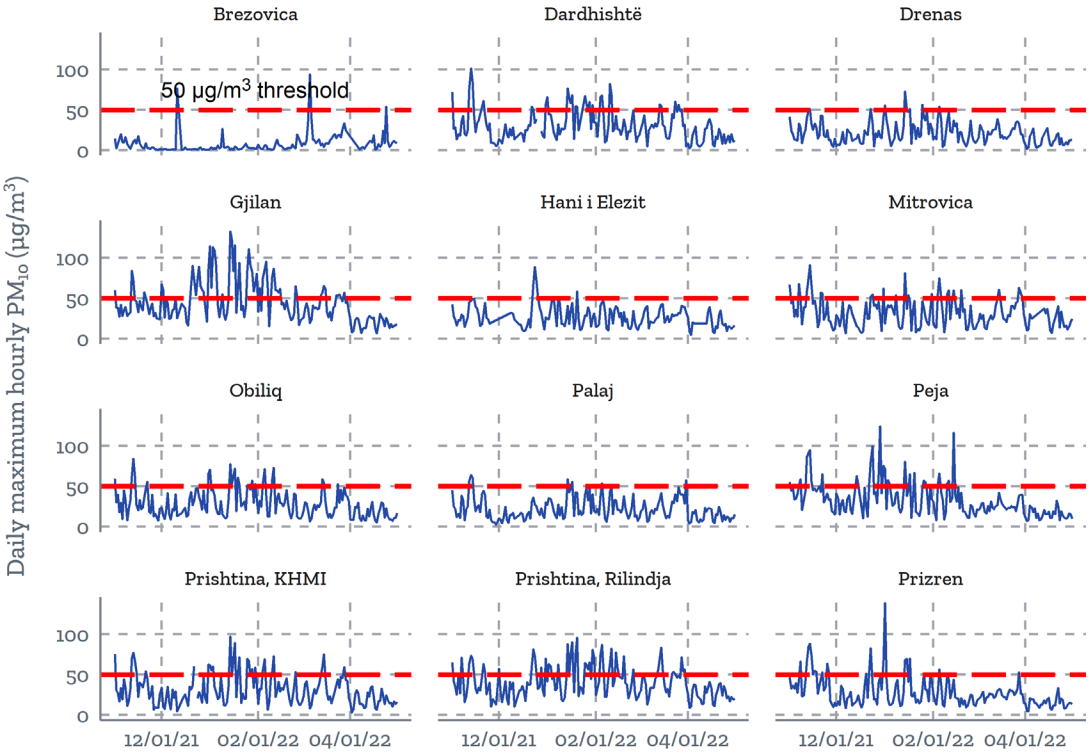
RQ4 is closely related to RQ3 and examines how the EDC activity has contributed to citizens using data to advocate for change. We will answer this question through interviews with CSOs and environmental activists, FGDs with users of the air quality data, and reviewing social media posts. We will rely on CAPT as an overarching research method through which we will use descriptive trends analysis to analyze and interpret quantitative data collected to answer this question. Key indicators for answering this question include documented use of air quality data to advocate for environmental or health improvements and barriers that are identified as major obstacles to effective citizen advocacy.

RQ5 aims to understand whether air quality data and outreach efforts affected citizens' behavior, including whether they use such information and act to reduce exposure to air pollution. We will use the air quality data to identify high pollution episodes to temporally anchor interviews and FGDs for gauging response measures people might be undertaking, such as staying indoors and avoiding exercise outside, purchasing air filters, or wearing masks outdoors.

The first two sub-questions of RQ6 focus on the quality and type of communication and engagement between organizations. Whereas RQ6a centers on the role that EDC may have played in easing tensions between civil society and the government, RQ6b addresses relationships among government agencies, particularly KEPA, NIPH, and KHMI. Implementation documents will be a valuable source for tracing how agencies collaborated on EDC activity components and to compile a record of examples where different organizations are jointly engaged in post-EDC activities, such as publishing reports, issuing press releases, or funding environmental initiatives. Such information will be helpful in triangulating conflicting accounts from different interviewees on matters such as whether cooperation has improved or the frequency of productive exchanges between civil society and government has increased. Knowing interviewees' tenure and role in these activities will help us determine whose accounts are most credible in cases where interviewees offer competing narratives. RQ6c will use correlation analysis to test KHMI's air quality forecast accuracy against measured values, the results from which will aid in interpreting data collected to answer RQ6a. We will measure the prevalence of monitor outages that result in missing air quality data and how often pollution exceeds predictions at the monitor level. For example, Figure III.6 displays daily mean PM_{10} values for the 12 fixed KHMI stations. Days with values at or

above the red horizontal line of 50 µg/m^3 constitute an exceedance under EU Directive 2008/50/IEC (EC 2022).

Figure III.6. Daily mean PM_{10} concentrations for the 12 air quality stations in KHMI's network for November 1, 2021 – May 1, 2022



Note: Mathematica calculations using data from KHMI.

Lastly, in RQ7 we will assess whether EDC improved the public's trust and understanding of GoK's role in environmental management. We will ask private citizens, CSO staff, and government officials whether they believe that public trust in the GoK has changed as a result of air quality-focused activities. We will assign more weight to the perspectives of individuals with longer professional engagement in issues of air quality and environmental policy in Kosovo, to triangulate conflicting accounts among stakeholders. We anticipate that academic researchers, who are presumably less biased on the topic of public trust than government employees or CSO staff, will also be able to provide input into whose perspectives are likely to be the most credible and reflect actual developments.

b. Study sample

Since the apps and information campaigns developed under the EDC activity were designed to be accessible to all Kosovans, the sample frame technically comprises all Kosovo residents. For the evaluation, we will primarily focus on individuals and organizations who have directly engaged with tools supported through the Threshold Program. We will define this group based on whether they have downloaded and used the air quality app or currently subscribe to the Facebook pages of relevant

environment-focused organizations. We will compare responses from this subgroup against a sample of individuals who have *not* engaged with any EDC materials or deliverables.[18]

We acknowledge that selection bias may affect which of these groups a person belongs to. For example, individuals with respiratory difficulties are more likely to access air quality-related materials, and so may differ substantively from the baseline population. As a result, any differences in perceptions or behavior between the two groups cannot be immediately interpreted as causally linked to EDC without following CAPT procedures to consider, investigate, and assess alternative explanations. Regardless, receiving input from both users and non-users of EDC deliverables is likely to provide richer insight into EDC's effects than only consulting individuals and organizations who have accessed air quality information.

c. Primary data collection

We will collect primary data for the EDC evaluation through both KIIs and FGDs. In Table III.13, we list potential interviewees or focus group populations, the key themes that would be addressed in those meetings, and the number of KIIs or FGDs we plan to conduct with each type of stakeholder. To incorporate additional or emerging perspectives, in some cases we identify more potential participants than we expect to need.

FGDs will supplement KIIs in our primary data collection by allowing individuals with similar relationships to the project to generate richer insights through discussion and interplay of experiences and ideas. Our proposed approach for recruiting FGD participants recognizes that the Kosovan public has had limited engagement with EDC-supported information outlets, according to recent ITT statistics shared by MCC. For example, the cumulative number of combined air quality app downloads from both the Google Play Store and iOS App Store is likely to be fewer than 2,000, whereas Kosovo's population is about 1.8 million people. Were we to adopt a population-wide random sampling strategy, it is highly unlikely that a sufficient number of respondents would be selected who have directly engaged with EDC outputs. Therefore, we propose to collaborate with KHMI to select FGD participants from among individuals who have interacted with one or more air quality platforms. One option for recruiting participants would be for KHMI to post on their Facebook page (~3,200 followers) an invitation to participate in a moderated discussion on air quality information and pollution perceptions.[19] To learn of the perspectives and experiences of individuals who have *not* engaged with EDC outputs like the app or the websites, we will use convenience sampling methods around the locations where FGDs will be conducted. Since EDC information campaigns included targeted efforts to specific groups, such as the elderly, individuals with chronic diseases, pregnant women, and Kosovo minorities, our recruitment efforts will ensure representation from these groups in our focus groups.

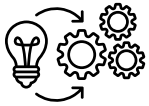
One of the motivations in conducting multiple interviews within a given stakeholder group, as well as across stakeholder groups, is to corroborate across sources to assess where there is consensus in perceptions, experience, or outcomes. Corroboration will be particularly helpful when diagnosing which factors mediated EDC's impact, and which factors were critical in determining why EDC did or did not

[18] To the best of our knowledge, no survey was conducted in the pre-treatment period that captures perceptions or behaviors in response to air pollution that could function as a baseline against which current values could be compared.

[19] Since KHMI and other government agencies do not have consent to share any contact information collected from individuals who have downloaded the relevant apps, this recruitment method enables individuals to directly contact us.

achieve its objectives. For example, consider a scenario in which multiple media outlets all receive the same communications training workshop, but only two outlets sustain regular broadcasts about air quality after the Threshold Program closes. Having multiple interviews, will enable us to discern the factors that led to different results for those two outlets from other workshop participants, whether because their staff had prior interest in environmental issues, they employed data analysts with the skills to analyze and interpret air quality data, or their outlet's leadership wishes to visibly champion public health causes. We believe this approach of multiple interviews will be particularly helpful for representatives from CSOs, the media, and the GoK to capture a fuller range of experiences that might be omitted if fewer interviews were pursued.

Table III.13. List of potential interviewees and FGD participants for EDC evaluation

Stakeholder type	Potential participants	Illustrative themes addressed	Target number
Key informant interviews			
MCC/MFK 	• MCC* • MFK*	• Project implementation and fidelity to initial plans • Perceived collaboration between government and civil society/media • Interpretation of trends observed in quantitative data sources	2–4 interviewees
GoK agencies 	• Kosovo Environmental Protection Agency (KEPA)* • Hydrometeorological Institute of Kosovo (KHMI)* • National Institute of Public Health (NIPH)* • Municipality of Pristina*	• Perceived relationship and level of engagement between government and civil society • Use of open data and relevant analyses in decision making • Inter-agency communication and collaboration	4–5
CSOs/NGOs 	• Institute for Development Policy • Peer Education Network • Health and Environment Alliance • Kosovo Environmental Program • Keep it Green • Kosovar Civil Society Foundation • Science for Change Movement	• Role of AQ data in social change advocacy • Experiences of underserved/vulnerable populations in accessing and using AQ information • Perceived relationship and level of engagement between government and civil society • Opportunities for and barriers to research/analyses conducted using AQ data	4–6
Other donors 	• Swedish Environmental Protection Agency • Japan International Cooperation Agency (JICA)* • Swedish International Development Cooperation Agency • World Bank	• Connections between EDC and other donor-funded projects • Perceptions of EDC effectiveness and key results • Role of donor coordination on EDC impacts	3
Implementers 	• NIRAS • Mott McDonald • Atmoterm • Vitech	• Implementation challenges and threats to project sustainability • Rationale for and results from actual AQ communications strategy	3

Stakeholder type	Potential participants	Illustrative themes addressed	Target number
Media 	• KALLXO.com/Internews Kosova • RTK • KTV* • T7	• User-friendliness of AQ communications • Determinants of how AQ information is communicated	3
Academics/ Researchers 	Kosovo-based researcher in field of air quality (name removed for this report)	• Quality of research using open data on AQ • Additional needs for using air pollution data to affect social change • Perceptions of public trust in and understanding of GoK's function related to environmental issues	2–3
Focus groups			
Mixed group consisting of users/visitors of KHMI AQ app or Facebook page **and** individuals who have not accessed AQ tools or services		• Usefulness and useability of available AQ information through phone app and portals • Actions undertaken to reduce pollution exposure • Trust in government and environmental policymaking	2 focus groups, with 6–10 participants each

Note: Asterisks signal potential participants who will be interviewed once for their experiences and perceptions across multiple TAG activities.

d. Secondary data collection

The EDC evaluation will draw upon many of the secondary data sources described in Chapter III, Section D.1. Table III.14 summarizes the secondary data sources we will rely on along with their purposes.

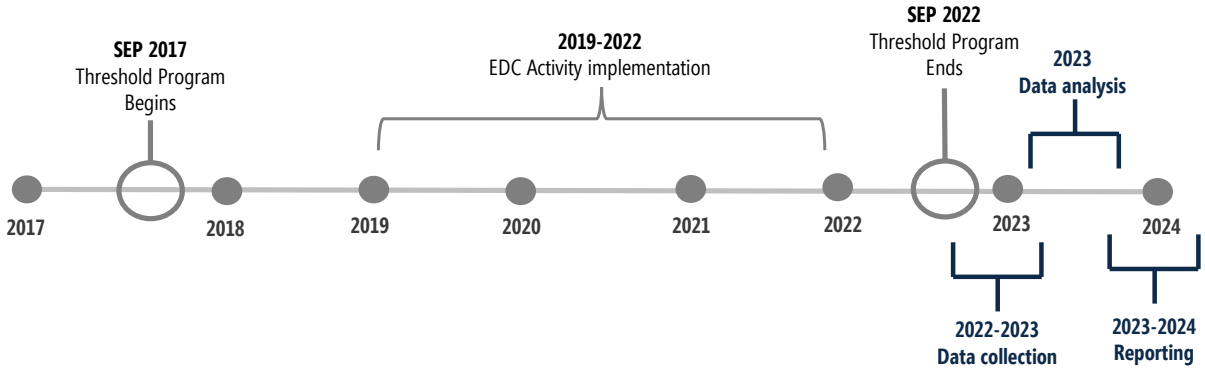
Table III.14. Overview of secondary data sources supporting the EDC evaluation

Data source	Contents	Purpose
Implementation documentation	Reports, meeting minutes, project summary documents, terms of references, indicator tracking tables, Threshold Program close-out reports, and any other documents encompassing EDC plans or actual implementation.	• Determining whether EDC achieved key outcomes • Identifying differences between plans and actual implementation
Data from TAG-supported portals and apps	Download counts and daily traffic volumes based on Google Analytics data or other software analytics output.	• Tracking changes over time in public use of AQ data • Assessing whether additional, relevant open data products have been released and their popularity
Google Trends data	Queries will be conducted in Albanian and English for pertinent search terms such as 'air pollution', 'air quality', and 'bad air.' The set of search terms will be finalized through consultation with MFK and CSOs.	• Detecting whether citizens' demand for AQ information has increased • Providing context for whether project achieved targeted outcomes
Facebook posts	Organizations/government agencies posting/reposting about AQ issues or data, public engagement levels with posts over time.	• Monitoring how organizations/individuals engage with AQ data and analyses • Discerning whether AQ data is supporting citizens' environmental advocacy efforts
KHMI AQ data (observations and forecasts)	Hourly station-level data for all AQ indicators across all sites since station monitoring began; daily countrywide 1-day and 3-day ahead forecasts.	• Quantifying trends in pollution exceedance rates • Estimating accuracy of AQ forecasts • Supporting qualitative data collection on individual behavioral change induced by AQ concerns
Usage statistics for KHMI site, NIPH micro-site, and AQ apps	Download counts and daily traffic volumes based on Google Analytics data or other software analytics output.	• Tracking changes over time in public use and dissemination of AQ data • Contextualizing engagement between civil society and the GoK

e. Timeline and exposure period

We will collect data in late 2022 through early 2023 (Figure III.7), allowing for approximately two years of exposure for some of the activity's key components. We note that in principle this should be a sufficient duration to observe project effects, especially because the activity promotes behavioral responses that can be immediately adopted and do not require long-term planning or investments. However, air quality information is more helpful on days of poor air conditions, and recent winters have been relatively mild; anecdotally, the mild winters have depressed the demand for air quality guidance and information, potentially limiting exposure to the availability of this information. We will analyze the primary and secondary data in Q1/Q2 of 2023 and share our final results in both reports and presentations in the period spanning Q4 of 2023 to Q1 of 2024.

Figure III.7. EDC data collection and analysis timeline



f. Analysis plan

For all KIIs and FGDs, we will develop guides and protocols in advance and submit them to MCC/MFK for review. The content of those guides and protocols will be informed by our review of the relevant literature on air quality monitoring systems and information communication, project documents, interviews conducted by Mathematica and by Moonshot Global during a June 2022 mission trip, and analysis of any available quantitative datasets prior to the KII/FGD. For all interviews, we will seek participants' consent to record the conversation. Individuals who are not comfortable speaking in English will be interviewed in Albanian, with interview transcripts or notes translated to English and individuals' statements anonymized. We will use respondent codes to ensure that any notable pull-out quotes or statements are linked to the correct respondent in line with best practices in reproducible research. All qualitative data sources—KIIs, FGDs, program documentation, and reports and analyses by CSOs and government agencies—will be coded in NVivo or another qualitative software to systematically extract key themes that will facilitate comprehensively answering all EDC-related research questions. Our processing of qualitative data will follow the qualitative thematic assessment and triangulation approach described earlier. Specifically, we will identify the incentives and levels of information access that each interviewee and or group of discussants has in order to assess the credibility of their assertions and triangulate their assertions, viewpoints, and experiences with those presented by other interviewees and discussants.

We have earlier outlined the quantitative datasets for which we will apply descriptive trends analyses, which include time-series data on Google Trends search intensity, website traffic, app installs and usage, and downloads of any open data on air quality available as part of EDC. Among those datasets, only the Google Trends data are available prior to the start of EDC activities, going back to 2016. We will use the full duration of data and compare the timings of above-average search activity against air quality measurements to understand the degree to which contemporaneous pollution conditions drive internet search patterns for air quality-related keywords. If we observe a lack of correlation between the two, suggesting that actual pollution levels are not a primary determinant of aggregate search behavior, we will identify episodes of above-average search activity and consult KII and FGD participants for their perceptions of other factors driving internet searches of air quality information. We will aim to schedule our FGDs during the winter to improve the recall accuracy of short-run defensive measures participants

may be engaging in to reduce air pollution exposure, such as staying indoors, running air purifiers inside their homes or workplaces, or wearing face masks while outside.[20]

For time-series data available on website traffic and app downloads, we will conduct descriptive trends analysis. As only post-treatment data are available, we will not be able to compare recent slopes and means against pre-treatment values, but we will be able to document whether usage patterns have persisted even after the end of EDC-supported activities like media broadcasts and the release of new videos on MFK's YouTube channel.

We will conduct basic statistical analyses of hourly and daily air quality data available from KHMI to assess air quality data availability and accuracy (RQ7c). We will tabulate the frequency of station outages, measured as the number of days with no publicly available measurements, and test for time trends. We will also report basic trends in air quality over time, including the frequency of days exceeding acceptable pollution concentration thresholds, to help contextualize changes in public perceptions of air quality and individuals' ability to engage in adaptive behavior. For example, were recent pollution conditions found to be substantively better than during the pre-treatment period, then such a trend could be a helpful explanation for reduced interest or concern about pollution exposure. We will conduct a correlation analysis by comparing real-time air quality data against 1-day and 3-day lead forecasts, for each indicator where real-time and forecast data are available. For each indicator and forecast lead time, we will estimate the Pearson correlation coefficient and calculate the root mean square error (RMSE), which is a measure of bias between the two datasets.

The CAPT framework described in Chapter III, Section C will be the overarching analytical approach used in answering EDC evaluation research questions. A key advantage of CAPT is its role in integrating insights from other research methods. For example, applying a descriptive trends analysis on website traffic statistics will uncover usage trends (e.g., growth, shrinkage, or stagnation), the timing of peak interest, and the types of content garnering the most engagement from website visitors. Such results can be directly interpreted through a CAPT lens or can be overlaid with other data sources such as interviews and FGDs to diagnose potential causes of such patterns and to highlight EDC's contribution to any observed dynamics.

g. Challenges and limitations

Below we mention a few challenges that are specific to evaluating the EDC activities.

- Services other than the platforms developed through EDC provide air quality information for Kosovo, such as the Up-to-date air quality data page hosted by the European Environment Agency and the AirNow Department of State site. By tracking user activity of EDC-funded information outlets, we will not capture all queries being made by Kosovans about air quality. As part of our qualitative data collection, we will ask stakeholders about the set of platforms and information sources on air quality that are commonly used to discern air quality as part of our contribution analysis whether air quality sources other than those supported under EDC may be driving behavior change or policy. While the availability of air quality information on non-EDC sources would prevent us from being able to collect accurate, country-wide statistics on information retrieval, we do not think this is a critical

[20] The health advisories and information contained in the AQ app provide more information about short-run measures individuals can perform than long-run adaptations like modifying the building envelope to reduce pollution intrusion into the home or switching to cleaner household heating solutions.

challenge to the evaluation. Our objective will be to observe and understand changes in usage and interest over time, which can be completed with a focus exclusively on EDC-supported platforms.

- There is no regularly conducted survey that addresses public perceptions of air quality or actions people take to minimize pollution exposure, which limits the scope for answering RQ6. Several surveys address an angle of air pollution perceptions or household actions, but they have been conducted only once and future rounds to form a pre-post comparison are not planned.[21] Since this is a retrospective evaluation without baseline data, baseline behaviors and attitudes can only be collected using recall questions, which are highly prone to bias or error.

3. KODC evaluation

The TAG Monitoring and Evaluation Plan indicates that the KODC activity's "theory of change is contingent upon the scale and adoption of the data transparency culture, and therefore the higher order outcomes may be modest" (MFK 2021, p. 21). Depending on the scale and use of KODC-produced apps and services, we will focus our analysis on the degree to which outputs, more than final outcomes, were achieved (particularly for RQ2). However, to offer insights on the success of the KODC activity, we propose a performance evaluation using a set of complementary analytical approaches, shown in Table III.15.

Table III.15. Overview of KODC evaluation design

Evaluation methodology	Key indicators	Proposed data sources
RQ1. Was the activity implemented according to plan (in terms of quantity and quality of outputs)?		
• Implementation analysis • Qualitative thematic analysis and triangulation	• Key indicators listed in the M&E plan • Stakeholders' reported use of data portals	• Implementation documentation (including implementation reports, ToR, work plans, program M&E Framework, Dig Data Challenge Data Guide, grant applications [both funded and nonfunded], internal reports generated by KODC management, MFK M&E data, reports submitted by grantees, and publicly available data on each grant) • KIIs
RQ2. Did the activity achieve its targeted outcomes, particularly its stated objective, in the timeframe and magnitude expected? Why or why not?		
• Implementation analysis • Descriptive trends analysis • Qualitative thematic analysis and triangulation	• Key program outcomes listed in the M&E plan achieved • Data used for decision making by citizens, government, and other relevant actors • Increased advocacy, collaboration and communication between government actors and civil society (by challenge type: air quality, energy, judicial, and labor force)	• Implementation documentation (see above) • Google Trends data • Data from online portals created by grantees • MCC Kosovo Scorecards[1] and ITT • KIIs and FGDs

[21] For example, the 2019—2020 Kosovo Multiple Indicator Cluster Survey and Roma, Ashkali and Egyptian Communities in Kosovo Multiple Indicator Cluster Survey asks respondents about the fuels and technologies they use for cooking. Although we would not be able to attribute any increase in the share of households adopting cleaner cooking fuels to EDC activities, simply being able to estimate any change would be helpful information in understanding the pace with which households have shifted to less polluting options in their daily activities.

Evaluation methodology	Key indicators	Proposed data sources
RQ3a. Is there any increase in the Government's use of analyses done by non-government entities, both analyses supported by MFK and those in general? Is there any evidence these types of analyses can factor into any policy decision-making? **RQ3b. Did publishing energy, labor force, air quality, and judicial data through relevant government websites result in increased analysis done by NGOs?** **RQ3c. Did KODC result in increased engagement between government and civil society/media?**		
• Descriptive trends analysis • Qualitative thematic analysis and triangulation • Political economy analysis	• Government agencies' use of NGO analysis • Government agencies' use of data to inform decision-making • NGOs' access and analysis of newly available data on government websites and portals • Improved relationship between GOK and civil society due to increased data access and government transparency • Government employees' perception of data as public good or resource to be shared • KODC data solutions that make public data more accessible to all citizens	• Implementation documentation • UNDP Public Pulse Survey data • Google Analytics data from websites and apps developed by KODC grantees – timing and usage statistics (if possible) • Google Trends data • KODC publicly available data including Dig Data Challenge-supported data publications • KIIs and FGDs
RQ4. How has the project contributed to citizens' use of open data in advocating for change?		
• Political economy analysis • Descriptive trends analysis • Qualitative thematic analysis and triangulation	• Perception of citizens' use of data to advocate for change • CSOs and citizens' use of data to advocate for change • Changes in organization of power, decision making, and economic resources among key actors (including changes in relative power of CSOs and citizens to advocate policy change)	• Social media content • KIIs and FGDs
RQ5. Did KODC contribute to increased trust and understanding of government's function?		
• Qualitative thematic analysis and triangulation • Political economy analysis	• Public perceptions of trust in government with respect to energy, environmental, labor, and judicial disclosures and policymaking • Changes in organization of power and decision making among key actors	• UNDP Public Pulse Survey data • KIIs and FGDs

[1] As with PAJI, we will refer to the data sources used for MCC Kosovo Scorecards, which ended in 2019. To the extent that more current data are available, we will collect them from the original sources.

a. Methodology

Our KODC evaluation will draw on each of the research methods listed in the first column of Table III.15. Across all research questions, we will process data from KIIs and FGDs with qualitative thematic analysis and triangulation to assess the credibility of assertions expressed in the data and the tensions and alignments in respondents' viewpoints and experiences.

We will apply an implementation analysis framework to answer RQ1 and RQ2 and uncover factors that contributed to any deviations in actual implementation against project plans. We will draw upon the project documentation that is available, including implementer progress reports, implementer annual reports, and minutes from donor coordination meetings as a starting point for follow-up discussions with MFK staff and implementers. For RQ2, which asks whether KODC achieved its targeted outcomes, we will draw insights from rich qualitative data from key informants and focus group discussants, including government stakeholders and civil society actors who are active in the rule of law space. To support our

descriptive trends analysis for RQ2, we will also draw on Google trends data, data from online portals created by KODC grantees, and data from the MCC scorecards and their sources.

Because the size, goals, and sectors of KODC grantees vary substantially, we will pair a portfolio-wide evaluation approach—one that analyzes and synthesizes the work of all grantees and their impacts—with case studies of four grantees with a variety of technical goals and levels of success to help answer RQs 3-5. These case studies (one from each DigData Challenge) will support a deeper dive into the implementation and contribution of grantees to open data collaboration and citizen understanding of government. We will request the help of MFK close-out staff[22] in selecting two grantees with exemplary open data analysis and products, and two grantees who struggled to deliver outputs and outcomes as planned. By contrasting these two pairs (and comparing grantee's sectors and sizes as well), we anticipate identifying key conditions for the success of open data challenge grantees. We also aim to illuminate any tensions grantees experienced in their work, such as opposing goals of addressing the immediate analytical needs of government agencies and developing innovative, more exploratory analytical products. We will supplement these case studies with FGDs that include all 22 grantees (see Table III.1).

RQ3 examines the degree to which the provision of energy, labor force, air quality, and judicial data has altered the ability of CSOs and the government to conduct new analyses based on that data, whether such analyses are informing policymaking, and whether KODC resulted in increased engagement between government and civil society or the private sector. While we will conduct descriptive trends analysis using Google Analytics and Trends data from websites and apps developed by KODC grantees, this RQ will primarily be answered through interviewing and conducting focus groups with government officials, representatives from open data organizations, and grantees themselves. We will apply political economy analysis to interpret changes in stakeholders' power, institutional arrangements, and interests, and will use qualitative thematic analysis and triangulation to help determine the strength of different viewpoints expressed in the qualitative data. We will also use our case studies of four grantees dive deeper into the conditions that support useful data products and strong relationships between CSOs and the government. Combining descriptive trends analysis with qualitative thematic analysis and triangulation and political economy analysis will help us trace the links between results stages and identify the contribution of KODC in light of concurrent initiatives to stimulate open data culture and CSO and government collaboration more broadly.

RQ4 is closely related to RQ3 and examines how the KODC activity has contributed to citizens' use of data to advocate for change. We will answer this question through interviews with CSOs and government officials, FGDs with grantees, and by reviewing social media posts. We will rely on political economy analysis, qualitative thematic analysis and triangulation, and descriptive trends analysis to process and interpret data collected to answer this question. Key indicators for answering this question include increased online citizen engagement with labor, air quality, environment, and judicial data to advocate for process improvements and barriers that are identified as major obstacles to effective citizen advocacy. As with RQ3, political economy analysis will allow us to catalog where the different stakeholder groups and institutions stand in terms of their level of support for KODC-supported tools and their level of influence, and descriptive trends analysis will help us identify changes in citizen engagement and advocacy. Combining the two analyses with qualitative thematic analysis and triangulation will be instrumental in interpreting the causes for any changes we find, including citizen engagement and trust outcomes

[22] We will also draw on information from the Moonshot Global/USAID Learning Mission that met with DigData grantees in June 2022

associated with KODC. We will also use our case studies of four grantees dive deeper into the conditions that help grantees develop data products that are useful for citizen advocacy for policy change.

Lastly, in RQ5 we will assess whether KODC improved the public's trust and understanding in the government's function. We will consult private citizens, CSO staff, and government officials to learn if they believe that public trust in the GoK has changed as a result of KODC activities. As with previous questions, we will process qualitative data with political economy analysis and with qualitative thematic analysis and triangulation and will deploy insights from our case studies of four grantees with various levels of success to illustrate how different conditions and approaches may contribute to public trust and understanding of government.

b. Study sample

Since many of the data products produced by grantees under the KODC activity were designed to be accessible to all Kosovans, the sample frame technically comprises all Kosovan citizens. For other areas of KODC, the sample frame is much smaller because some grantees targeted such groups as high school youth or a specific government agency. For the evaluation, we will gather qualitative data collection on individuals and organizations who have directly engaged with tools supported through the Threshold Program. We define this group by identifying organizations and individuals who are involved in open data work in Kosovo or who have used KODC grantees' products as private citizens, civil society actors, or government officials. We will also collect qualitative data from individuals and organizations who are not directly involved with the Threshold Program but who can offer insights on the perceptions and experiences of government officials, CSOs, media, and other groups as they relate to open data access, analysis, and relationships (see Table III.16).

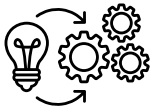
c. Primary data collection

We will collect primary data for the KODC evaluation through both KIIs and FGDs. In Table III.15, we list potential interviewees or focus group populations, the key themes that would be addressed in those meetings, and the number of KIIs or FGDs we plan to conduct with each type of stakeholder. To incorporate additional or emerging perspectives, in some cases we identify more potential participants than we expect to need.

As the KODC activity engages a wide variety of stakeholder institutions and groups, we pursue a broad qualitative primary data collection strategy targeting between 21 and 25 interviewees and 4 focus groups comprising up to 24 individuals. One of the motivations in conducting multiple interviews within a given stakeholder group, as well as across stakeholder groups, is to corroborate across sources to assess where there is consensus in perceptions, experience, or outcomes. Corroboration will be particularly helpful when diagnosing which factors mediated KODC's impact, and which factors were critical in determining why grantees achieved (or did not) their objectives. In addition to delivering corroboration, we believe interviewing multiple interviewees in one stakeholder group can allow us to reach saturation—a point at which we have collected all major (even diverging) viewpoints and each additional interview yields diminishing marginal value in data.

Table III.16. List of potential interviewees and FGDs participants for KODC evaluation

Stakeholder type	Potential participants	Illustrative themes addressed	Target number
Key informant interviews			
MCC/MFK 	• MCC* • MFK*	• Project implementation and fidelity to initial plans • Perceived collaboration between government and civil society/private sector • Interpretation of trends observed in quantitative data sources	2–4 interviewees
GoK agencies 	• Ministry of Finance, Labor, and Transfers; Kosovo Environmental Protection Agency (KEPA)*; Kosovo Hydrometeorological Institute (KHMI) *; National Institute of Public Health (NIPH) *; Kosovo Judicial Council (KJC) *; Kosovo Prosecutorial Council (KPC)*; Agency of Statistics; Energy Regulatory Office (ERO); Municipality of Pristina*; KOSTT j.s.c, Agency of Gender Equality*	• Use of open data and relevant analyses and products from CSO/private sector in GoK decision-making • Perceived effects of KODC on data engagement of the public and on empowerment of women and youth • Perceived relationship and level of engagement between government and civil society	5–6
Professional groups 	• Kosovo Economic Chamber	• Perceptions of open data culture and innovations in Kosovo • Information on relationship between open data and development	1
CSOs/NGOs 	• Open Data Kosovo • Kosovo Open Society Foundation • Kosovo's Women's Network	• Status of open data in Kosovo and role of open data in social change advocacy • Perceived effects of KODC on data engagement of the public and on empowerment of women and youth • Perceived relationship and level of engagement between government and civil society	3
Other donors 	• USAID • JICA*	• Connections between KODC and other donor-funded projects • Perceptions of KODC effectiveness and key results • Role of donor coordination on KODC impacts	2

Stakeholder type	Potential participants	Illustrative themes addressed	Target number
Non-grantee private sector 	• Kosovo Power Solutions	• Perceptions of availability of and demand for energy data for ongoing analysis	1
Grantees 	• Subset of four grantees with a variety of technical goals and levels of success, to be determined in consultation with MFK close-out staff and with information from the Moonshot Global/USAID Learning Mission that met with DigData grantees in June 2022	• Usefulness and useability of datasets available for the challenges • Experience with challenge application and grant award processes • Experience with developing data analyses and products • Perceptions of partnerships formed between GOK and civil society/private sector and associated collaboration; sustainability of those outcomes • Perceptions of citizen engagement with grantees' data analyses and products • Perceptions of grantee capacity changes during KODC and sustainability of grantee organizations and data products	4
Media 	• Radio TV of Kosovo* • Radio TV 21* • KTV* • Online written media sources*	• Use of KODC grantees' analyses or data products for reporting • Perception of CSO, private sector, and citizen engagement with government through data products	2–3
Academics/ Researchers 	Kosovo-based researcher in field of government transparency and civil society (name removed for this report)	• Perceptions of partnerships formed between GOK and civil society/private sector and associated collaboration and communication • Perceptions of public trust in and understanding of GOK's function	1

Stakeholder type	Potential participants	Illustrative themes addressed	Target number
Focus groups			
	Mixed groups of leaders or representatives of the 22 KODC DigData Challenge grantees, divided into groups by challenge area (labor, judicial, energy, and air quality)	• Usefulness and useability of datasets available for the challenges • Experience with challenge application and grant award processes • Experience with developing data analyses and products • Perceptions of partnerships formed between GOK and civil society/media and associated collaboration; sustainability of those outcomes • Perceptions of citizen engagement with grantees' data analyses and products • Perceptions of grantee capacity changes during KODC and sustainability of grantee organizations and data products	4 focus groups, with 4–7 participants each

Note: CSO = Civil society organization; MFK = Millennium Foundation Kosovo; NGO = Non-governmental organization. Asterisks signal potential participants who will be interviewed once for their experiences and perceptions across multiple TAG activities.

As indicated above, we will conduct case studies of four grantees with a variety of technical goals and levels of success to help illuminate the implementation experience of the organizations and the conditions under which they achieve the best outcomes in terms of data communication and collaboration. For those four organizations, we will conduct KIIs with their leaders and invite them to join our focus groups with representatives of all grantees, as well. We will use these four FGDs to efficiently gather rich, nuanced information from grantees who share the same DigData Challenge, but whose backgrounds and experiences with the competition and partners are distinct.

d. Secondary data collection

The KODC evaluation will draw upon many of the secondary data sources described in Chapter III, Section D.1. Table III.17 summarizes the secondary data sources we will rely on along with their purposes.

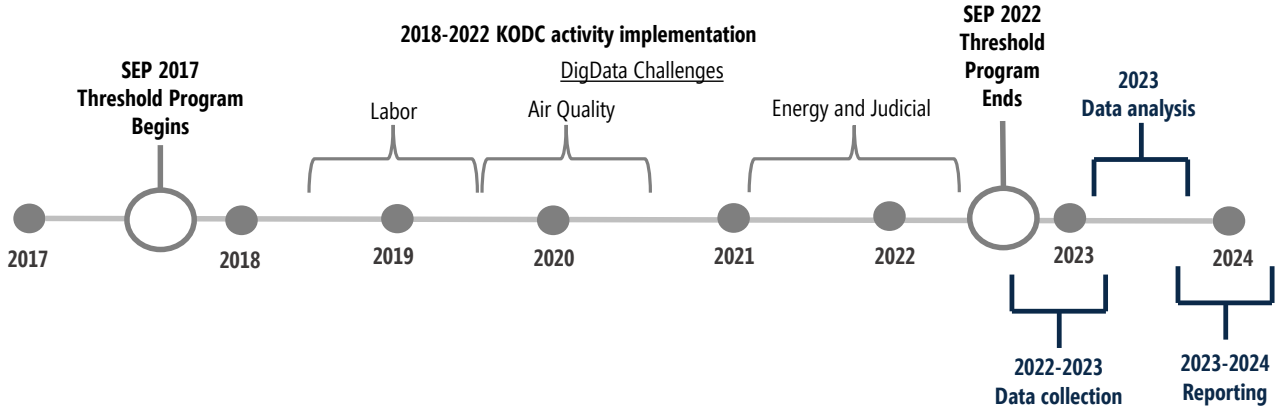
Table III.17. Overview of secondary data sources supporting the KODC evaluation

Data source	Contents	Purpose
Implementation documentation	Reports, meeting minutes, project summary documents, terms of references, KODC grantee concept notes, milestone reports, indicator tracking tables, Threshold Program close-out reports, and any other documents encompassing KODC plans or actual implementation.	• Determining whether KODC achieved key outputs and outcomes • Identifying differences between plans and actual implementation
Kosovo Public Pulse Surveys	Measurements of citizen trust in key government institutions over time	• Contextualizing qualitative data with longer-term trends in citizen perspectives
Google Trends data	Queries will be conducted in Albanian and English for pertinent search terms such as 'open government data'. The set of search terms will be finalized through consultation with MFK and CSOs.	• Detecting whether citizens' demand for open government data has increased • Providing context for whether project achieved targeted outcomes
Facebook posts	Organizations/government agencies posting/reposting about open government data portals and analyses, public engagement levels with posts over time.	• Monitoring how organizations/individuals engage with open government data and analyses • Discerning whether open government data is supporting citizens' advocacy efforts
General usage statistics for grantees' apps and websites	Download counts and daily traffic volumes based on Google Analytics data or other software analytics output.	• Tracking changes over time in public interest in open government data in DigData Challenge areas • Contextualizing engagement between civil society and the GoK

e. Timeline and exposure period

The implementation of KODC varied by DigData challenge area, with the Labor Challenge taking place 2018-2019, Air Quality in 2019, and Energy and Judicial Challenges taking place in 2021–2022 (Figure III.8). Given this staggered approach, our data collection in Q3 and Q4 of 2022 will gather insights on programs that concluded up to three years ago and as late as August 2022.

Figure III.8. KODC data collection and analysis timeline



As indicated in the evaluability assessment (Chapter III, Section B), we expect that changes in participants' behaviors (such as citizens using new analyses to advocate for change with government) are dependent on changes in those participants' capabilities, opportunities, and motivations, which may take time to accrue. At the same time, a long lag between the closure of the DigData Labor Challenge and our data collection may pose issues in terms of stakeholder recall and grantee turnover. Given that issue, we may need to conduct more intensive outreach with labor grantees and the government agency that provided the original dataset and re-share data products they developed and used to jog their memories. For Air Quality, Energy, and Judicial Challenges, we will ask grantees and associated government agencies for their thoughts on their expectations for data product maintenance and use over the next two years, to approximate data we can collect with Labor Challenge grantees and partners. We will analyze the primary and secondary data in Q1/Q2 of 2023 and share our final results in reports and presentations in the period spanning Q4 of 2023 to Q1 of 2024.

f. Analysis plan

For all KIIs and FGDs, we will develop guides and protocols in advance and submit them to MCC/MFK for review. The content of those guides and protocols will be informed by our review of the relevant literature on open judicial data systems and information communication, project documents, interviews conducted by Mathematica and by Moonshot Global during a June 2022 mission trip, and analysis of any available quantitative datasets prior to the KII/FGD. For all interviews and focus groups, we will seek participant consent to record the conversation. Individuals who are not comfortable speaking in English will be interviewed in Albanian, with interview transcripts or notes translated to English and individuals' statements anonymized. We will use respondent codes to ensure that any notable pull-out quotes or statements are linked to the correct respondent in line with best practices in reproducible research.

All qualitative data sources—KIIs, FGDs, program documentation, and reports and analyses by CSOs and government agencies—will be coded in NVivo or another qualitative software to systematically extract key themes that will facilitate comprehensively answering all KODC-related research questions. Our processing of qualitative data will follow the qualitative thematic assessment and triangulation approach described earlier. Specifically, we will identify the incentives and levels of information access that each interviewee and or group of discussants has in order to assess the credibility of their assertions and triangulate their assertions, viewpoints, and experiences with those presented by other interviewees and discussants. We will use political economy analysis to understand relationships between civil society and private sector actors with government agencies and the broader political forces operating which may affect collaboration on open data analyses and displays of citizen advocacy.

We have earlier outlined the quantitative datasets for which we will apply descriptive trends analyses, which includes time-series data on Google Trends search intensity and app and website traffic for KODC grantees' data products. Among those datasets, only the Google Trends data are available prior to the start of PAJI activities. We will use the full duration of available data and compare the timings of above-average search activity against important events in the open data ecosystem (such as MFK's call for DigData applications or release of YouTube videos, or grantee' publication of new platforms) to understand the degree to which contemporaneous media coverage and analyses drive internet search patterns for open data-related keywords. If we observe a lack of association between the two, suggesting that the visible actions of government, media, and civil society are not a primary determinant of aggregate search behavior, we will identify episodes of above-average search activity and consult KII and FGD participants for their perceptions of other factors driving internet searches for open government data products.

Finally, we will use data from the UNDP Public Pulse Surveys related to citizen satisfaction in their government institution to contextualize trends in media coverage, trends in social media engagement and website traffic, and qualitative insights we gain on open data use for advocacy.

The CAPT framework described in Chapter III, Section C will be the overarching analytical approach used in answering KODC evaluation research questions. A key advantage of CAPT is its role in integrating insights from other research methods. For example, applying a descriptive trends analysis on website traffic statistics will uncover usage trends (e.g., growth, shrinkage, or stagnation), the timing of peak interest, and the types of content garnering the most engagement from social media users. Such results can be directly interpreted through a CAPT lens or can be overlaid with other data sources such as interviews and FGDs to diagnose potential causes of such patterns and to highlight KODC contribution to any observed dynamics in collaboration and advocacy based on open government data. To further strengthen our contribution analysis, we will interview the four grantees previously selected for our case study[23] to gather details on how those organizations conceptualized their projects, developed programs and products, collaborated with government institutions and other partners, and disseminated their work. We will pair this primary qualitative information with details from all documents available on the four selected grantees to form a clear picture of the experience and contribution of DigData grantees.

g. Challenges and limitations

Particularly for the DigData Labor Challenge, which began in 2018, staff turnover among grantees or the dissolution of certain grantee organizations may impede qualitative data collection. To increase the likelihood that we can collect data from individuals involved in the implementation of DigData grants, we will work with MFK close-out staff to identify the latest contacts for each grantee and initiate conversations with those organizations and companies as soon as possible.

As indicated in Chapter III, Section B, the KODC activity has weaker evaluability than PAJI and EDC. KODC lacks comprehensive project justification and planning at the grantee and activity level, particularly in terms of identification of risks, assumptions, and mitigation strategies, clarity on project beneficiaries, and measurement and monitoring systems. This challenge may limit the evaluation's ability to make clear assessments of KODC's outcomes such as *increased collaboration and communication between GoK and civil society/private sector.* To address this challenge, we will:

1. Closely review notes and findings from the June 2022 Moonshot Global/USAID Learning Mission to develop an initial understanding of KODC's (and grantees') risks and assumptions;
2. Work with MFK close-out staff to ensure we have access to all documents and data sources that could support evaluability; and
3. Use contribution analysis with process tracing with a subset of 4 selected grantees to organize and contextualize the insights from each research method and strengthen our understanding of the activity's impacts.

[23] As mentioned in the KODC primary data collection section of this chapter, we will conduct interviews with four grantees selected in consultation with MFK that have a variety of technical goals and levels of success to help illuminate the implementation experience of the organizations and the conditions under which they achieve the best outcomes in terms of data communication and collaboration. These four organizations will also be included in focus groups with all grantees.

4. Cross-cutting evaluation

We propose a cross-cutting performance evaluation that examines TAG as a whole which we summarize in Table III.18.

Table III.18. Overview of cross-cutting TAG evaluation design

Evaluation Methodology	Key Indicators	Proposed Data Sources
RQ1. Did the program achieve its targeted outcomes, particularly its stated objective, in the timeframe and magnitude expected? Why or why not?		
• Implementation analysis • Descriptive trends analysis • Qualitative thematic analysis and triangulation	• Key program outcomes listed in the M&E plan achieved • Data used for decision making by citizens, health advisories, government, and other relevant actors • Increased advocacy, collaboration and communication between government actors and civil society	• Implementation documentation • KIIs • Google Trends data • Data from online portals created by TAG Project • MCC Kosovo Scorecards and ITT • Social media content
RQ2. Do the results of the program justify the allocation of resources towards it?		
• Qualitative thematic analysis and triangulation	• Stakeholder impressions of project's impact on future economic prosperity in Kosovo • Perceptions about the benefits of open data and the role that TAG played in furthering open data efforts	• KIIs, FGDs • MCC ITT • Financial data for each TAG activity • Social media content
RQ3. Is there a change in government employees' perceptions of government data as a public good or as a resource to be shared? If yes, how are government employees sharing data with the public (open data?, website?, reports?)? If no, why?		
• Qualitative thematic analysis and triangulation	• Perceptions and experiences of government employees on data sharing and disclosure • Channels through which government data is shared	• KIIs

a. Methodology

Similar to the activity-specific evaluations, we will use implementation analysis, descriptive trends analysis, and CAPT to address the overarching TAG-wide research questions listed in Table III.16. In contrast with the activity-specific evaluations, this evaluation will feature an additional layer of synthesis to distill common features of success or difficulty that occurred across the three activities.

b. Study sample

The study sample for the cross-cutting evaluation is the union of study samples from the three activity evaluations.

c. Primary data collection

The cross-cutting evaluation will leverage all interviews conducted in support of the three activity-level evaluations. To assess linkages and synergies across activities we will ask questions elevated to themes like "open data culture" and "transparency in governance" which transcend the provenance of any individual activity. Such questions will aim to prompt discussions about whether components of the three activities interacted (or did not interact) and the role such interaction – in contrast to activities in isolation – had in transforming government processes or the attitudes of government employees. This will be

particularly relevant in answering RQ2 about whether the program generated results that are cost-justified. Because project benefits may arise beyond the scope of any individual activity, adopting a Kosovo-wide focus will facilitate capturing broader transformations in the collection, reporting, and use of data outside the sectors directly targeted through TAG.

d. Secondary data

We will primarily rely on the secondary data sources described in the activity-level evaluations. Data sources specific to addressing cross-cutting features will include broader search terms in Google Trends queries, like "open data" or "open data platforms", as well as collecting visit and download statistics from the Kosovo Agency of Statistics who is the main proprietor of open datasets for the GoK.

e. Timeline and exposure period

We will collect any datasets specific to answering these research questions when data are collected for the other evaluations. We will conduct data analysis in Q1/Q2-2023 and share our findings in late-2023. The exposure period will be defined according to whether treatment begins when project activities began, or when project activities ended. If the latter, then the exposure period for overarching TAG activities will be no greater than six months given delays in the PAJI Activity implementation. If the former, then the exposure period would exceed two years.

f. Analysis plan

As with the preceding evaluations, CAPT will be the primary tool with which we process data and synthesize our findings. For documents and interview segments focusing on TAG synergies and cross-activity linkages, we will develop a coding scheme that enables us to separate out project-specific findings from activity-specific ones. As inputs into the CAPT, we will perform a descriptive trends analysis on Google Trends search intensities to identify if there are spikes in activity that might be connected to specific TAG milestones or events. To answer RQ2 about the program's cost-effectiveness, we will inquire with stakeholders about the benefits they perceive to have resulted from TAG to determine whether that complete set of benefits exceeds the project's expenditure of approximately $8 million. We will validate each purported benefit stream shared by interviewees by consulting other stakeholders and gaining multiple peoples' perception about the probability of such outcomes happening in the absence of TAG activities. Through contribution analysis, we will be able to make more definitive statements about TAG's role in observed outcomes that could be priced as inputs to the cost-effectiveness analysis. Because many of the project's target outcomes – perceptions of government performance, trust in government, support for open data, and improved relations between civil society and GoK – are both difficult to accurately measure let alone convert into monetary values, we are not pursuing a quantitative approach when examining project cost-effectiveness.

F. Evaluation risks and mitigation strategies

Although we believe there are few risks that could affect the evaluation, in this section we discuss those we foresee as the most likely and describe the primary mitigation strategies we will adopt to minimize risk impacts.

Securing access to website traffic and download data. One of the challenges with collecting administrative and secondary data after the Kosovo Threshold Program ends is that MFK staff will not be available to facilitate data-sharing agreements and, if necessary, nudge data owners who are unresponsive

or slow in providing data needed for the evaluation. To minimize the risk of delayed data transfer to the evaluation team, we reached out to key government agencies and implementers during an in-country mission trip to discuss our data needs and develop rapport with those organizations. We have in-country team members who will be able to make in-person visits, if needed, to increase the chances we receive all the datasets necessary for a successful evaluation.

Turnover in key agencies and organizations involved in the project. With TAG activities beginning as early as 2017, at least some government officials, MFK personnel, and implementers involved with the project are likely to have left their positions since the project began. As a result, some interviewees may be unable to share a perspective of the early phases of TAG, or the environment before TAG was launched.[24] To minimize this issue, we will schedule interviews with MFK staff, implementers, and critical government staff in the earliest portion of our data collection process to reduce the risk of future turnover affecting interviewees' availability. We will also identify other interviewees with longer tenures at key institutions to supplement or substitute for individuals named by organizations who were only recently hired or assigned to a TAG-relevant portfolio.

[24] Turnover among CSO and media staff poses a smaller risk to collecting quality data, because interviews with those stakeholders will focus on current and ongoing issues, like trust in government, data collaboration, and use of CSO analyses.

IV. Evaluation administration and management

In this chapter, we discuss administrative issues relevant to managing the evaluation and present a timeline of the evaluation activities.

A. Summary of institutional review board requirements and clearances

Mathematica is committed to protecting the rights and welfare of human subjects and will prepare and submit an application for approval of the research and data collection plans to an institutional review board (IRB) registered with the Office for Human Research Protections within the U.S. Department of Health and Human Services. We intend to use Health Media Lab as our IRB. We will submit the required documents, including a research protocol providing details of the study and data collection activity, copies of all data collection instruments, and a completed IRB questionnaire that summarizes the key elements of the research protocol and plans for protecting participants' confidentiality. The data collection instruments that we will prepare and share with the IRB will include consent statements that guarantee the confidentiality of respondents to the extent possible.

We will provide evidence of the IRB approval to MCC. If data collection instruments change substantially from those that the IRB approved, then we will reapply for review. Small changes to the instruments (such as rewording of questions, reordering of questions, or editing changes) do not require reapplication, but the finalized instruments must be submitted to the IRB for documentation. We will submit the instruments for review in both English and Albanian as applicable.

B. Data protection

Mathematica and our local consultants will ensure confidentiality of all respondents, including confidentiality of participation in the data collection, confidentiality of personally identifiable information, and other sensitive data. When quantitative or qualitative data are collected, we will ensure the safe handling and transfer of electronic files. Electronic data files will be shared with Mathematica using a secure file transfer system, such as a file transfer protocol, file exchange website (FX site), or a SharePoint site. All files with sensitive information, including those for secondary data analyses and document review, will be stored in a designated encrypted project folder, which is secured with AES 256-bit encryption. Data files will be accessible only to project team members who clean or analyze the data. All project team members have signed a nondisclosure agreement pertaining to confidential information.

C. Preparing data files for access, privacy, and documentation

Primary data collected for this evaluation will be qualitative. As qualitative data are inherently susceptible to identification, and deidentification normally renders them unusable, we will not prepare these for delivery.

D. Dissemination plan

In addition to writing a final report of our evaluation, the Mathematica team will provide input to an evaluation brief for the final results that will include methods and key results. We will also present the final evaluation findings in person both to MCC and to technical and non-technical project stakeholders and policymakers in Kosovo. Furthermore, we will participate in any other MCC-financed dissemination and training events related to the findings from the final evaluation reports. We expect the broader

research community to have strong interest in the findings from the evaluation. To facilitate wider dissemination of findings and lessons, we will collaborate with MCC and other stakeholders to identify additional forums—conferences, workshops, and publications—to disseminate the results and encourage other donors and implementers to integrate the findings into their programming.

E. Evaluation team's roles and responsibilities

Our team has extensive experience and expertise in evaluation design, data collection, and analysis to meet MCC's evaluation needs. **Mr. Matt Sloan** oversees the project team and provides technical leadership. He is responsible for managing the evaluation team, leading the design and implementation of the evaluation, and overseeing data collection efforts. Mr. Matt Sloan also monitors the project's budget and schedule, and he manages communication with MCC, local partners, and other stakeholders. **Dr. Anthony Louis D'Agostino** works alongside Mr. Sloan to oversee the technical aspects of the evaluation, including the development of the recruitment strategy, instruments, and coding scheme, analysis, and reporting for the process evaluation and case studies. **Mr. Josh Meuth Alldredge** works with Mr. Sloan and Dr. D'Agostino to develop both the qualitative data collection instruments, coordinate with local consultants on data collection, and contribute to the analyses and reporting on all four evaluation components. **Ms. Hailey Hannigan** and **Ms. Ksenia Miliutinskaia**, research assistants, provide support for instrumentation development, data collection management, data analysis, reporting, and overall project coordination. **Mr. Randall Blair** provides quality assurance reviews for all key deliverables for this project.

Mathematica is working closely with a team of expert consultants with a unique combination of the qualifications and skills needed to conduct a successful evaluation of TAG. **Dr. Ariana Qosaj-Mustafa** is a researcher for WI-HER and has deep experience with the Kosovar judiciary system. **Ms. Rinora Gojani** serves as the environmental and research and evaluation expert. Our consultants will also arrange site visits of Mathematica personnel, keep our team apprised of program developments, assist with interviews and in-country data collection, and assist with communications with relevant stakeholders after the program closes.

F. Evaluation timeline and reporting schedule

The evaluation activities presented in Table VI.1 below include one round of data collection in which Mathematica will procure the data for all three TAG program activities. The qualitative data collection will be conducted in the third quarter of 2022 and will continue through the first quarter of 2023. In the second quarter of 2023, we will conduct the data cleaning and analysis and expect to submit our first draft report concurrently. We expect to submit the final evaluation report by the third quarter of 2023, incorporating feedback from stakeholders after the presentation of the draft report.

Table IV.1. Evaluation timeline

Task	Task/Deliverable Name	Estimated Due Date
Develop Evaluation Design Report	Draft Evaluation Design Report	8/26/2022
	MCC EMC presentation[a]	9/2023
	Final Evaluation Design Report	9/16/2022
	Draft and Final Covid Risk Mitigation Plan	9/16/2022
	Validate evaluation Metadata for Evaluation Catalog entry	9/1/2022
Develop Evaluation Materials	Draft and finalize English guides and protocols (for KIIs and FGD)	For MFK/implementers: 8/26/2022 For other stakeholders: 10/7/2022
	Translation, summary of pre-test test, written review of back-translation	For MFK/implementers: 9/14/2022 For other stakeholders: 10/21/2022
Undertake Data Collection	IRB approval	9/14/2022
	Written summary of quality control checks	9/30/2022
	Conduct interviews with MFK and implementers	9/26/2022 - 10/24/2022
	Conduct data collection for EDC and KODC (KIIs, FGDs, administrative data)	11/4/2022 - 12/30/2022
	Conduct data collection for PAJI (to allow longer exposure time)	1/2/2022 - 3/31/2023
Develop Final Report	Review and synthesize administrative data	3/31/2023
	Conduct qual data analysis	4/31/2023
	Conduct quant data analysis	4/31/2023
	Draft Evaluation Report	6/30/2023
	Data Package as per MCC Transparent, Reproducible, and Ethical Data and Documentation Guidance	7/31/2023
	Final Evaluation Report; Public Statement(s) of Difference/Support	8/31/23
	Executive Summary of Final Report in Albanian	9/30/23
	Evaluation Brief content in English and Albanian	9/30/23
Disseminate Final Report	Prepare metadata	9/30/23
	Presentation materials; validation of evaluation Metadata	9/30/23
	MCC presentation	10/31/2023
	Stakeholder presentation	10/31/2023

[a] Exact date to be determined.

This page has been left blank for double-sided copying.

References

Aguilar-Gomez, Sandra, Holt Dwyer, Joshua S. G. Zivin, and Matthew J. Neidell. "This is Air: The 'Non-Health' Effects of Air Pollution." NBER Working Paper Series, 2022.

Kosovo Environmental Protection Agency. "Annual Report on the State of the Environment for 2020." 2021. Available at https://ammk-rks.net/assets/cms/uploads/files/Publikime-raporte/ANG_Web.pdf.

Babamusta, Ermira, "Political Trust in Kosovo: Exploring Cultural and Institutional Dynamics." Doctoral Dissertation. Morgantown, WV: West Virginia University, 2019. https://researchrepository.wvu.edu/etd/3918.

Bajpai, Ranji, and C. Bernard Myers. "Enhancing Government Effectiveness and Transparency: The Fight Against Corruption." Washington, DC: World Bank, International Bank for Reconstruction and Development, September 2020. Available at https://documents1.worldbank.org/curated/en/235541600116631094/pdf/Enhancing-Government-Effectiveness-and-Transparency-The-Fight-Against-Corruption.pdf.

Barkjohn, Karoline K., Brett Gantt, and Andrea L. Clements. "Development and Application of a United States-Wide Correction for PM2.5 Data Collected with the PurpleAir Sensor." *Atmospheric Measurement Techniques*, vol. 14, no. 6, 2021, pp. 4617-4637.

Barwick, Panle J., Shanjun Li, Liguo Lin, and Eric Zuo. "From Fog to Smog: The Value of Pollution Information." *NBER Working Paper Series*, Working Paper 26541, 2019.

Bernauer, Thomas and Vally Koubi. "Effects of Political Institutions on Air Quality." *Ecological Economics*, vol. 68, no. 5, 2009, pp. 1355-1365.

Bi, Jianzhao, Avani Wildani, Howard H. Chang, and Yang Liu. "Incorporating Low-Cost Sensor Measurements into High-Resolution PM2.5 Modeling at a Large Spatial Scale." *Environmental Science & Technology*, vol. 54, no. 4, 2020, pp. 2152-2162.

Bickerstaff, Karen and Gordon Walker. "Public Understandings of Air Pollution: The 'localisation' of Environmental Risk." *Global Environmental Change*, vol. 11, no. 2, 2001, pp. 133-145.

Boso, Àlex, Aner Martínez, Marcelo Somos, Boris Álvarez, Constanza Avedaño, and Álvaro Hofflinger. "No Country for Old Men. Assessing Socio-Spatial Relationships between Air Quality Perceptions and Exposures in Southern Chile." *Applied Spatial Analysis and Policy*, 2022, pp. 1-18.

Chay, K. Y. and M. Greenstone. "The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession." *The Quarterly Journal of Economics*, vol. 118, no. 3, 2003, pp. 1121-1167.

Chen, Shuai, Paulina Oliva, and Peng Zhang. "Air Pollution and Mental Health: Evidence from China." *NBER Working Paper Series*, 2018.

Cieślik, Andrzej and Łukasz Goczek. "Control of corruption, international investment, and economic growth – Evidence from panel data." *World Development,* vol. 103, March 2018, pp. 323-335. https://doi.org/10.1016/j.worlddev.2017.10.028.

GBD 2019 Risk Factors Collaborators. "Global Burden of 87 Risk Factors in 204 Countries and Territories, 1990–2019: A Systematic Analysis for the Global Burden of Disease Study 2019." *The Lancet*, vol. 396, no. 10258, 2020, pp. 1223-1249.

Cori, Liliana, Gabriele Donzelli, Francesca Gorini, Fabrizio Bianchi, and Olivia Curzio. "Risk Perception of Air Pollution: A Systematic Review Focused on Particulate Matter Exposure." *International Journal of Environmental Research and Public Health*, vol. 17, no. 17, 2020, pp. 6424.

Dechezleprêtre, Antoine, Nicholas Rivers, and Balazs Stadler. "The Economic Cost of Air Pollution: Evidence from Europe." *OECD Economics Department Working Papers*, no. 1584, 2019. https://doi.org/10.1787/56119490-en.

Delmas, Magali A. and Aanchal Kohli. "Can Apps make Air Pollution Visible? User Engagement with Air Quality Information." *SSRN Electronic Journal*, March 2019. http://dx.doi.org/10.2139/ssrn.3353132.

Dahmani, Maher and Sophie Vermeille. "Open Data for Case Law: A Digital Republic for Predictable and Attractive Legal Rules." *SSRN*, September 2017. http://dx.doi.org/10.2139/ssrn.308484.

Data.org. "Reporting Back to Our Community: Insights and Learning from the data.org Inclusive Growth and Recovery Challenge." New York, NY: Data.org. Available at https://data.org/reports/learning-insights-challenge-report/.

Domagala, Natalia. "What technology and open data can do for women in Kosovo: A critical assessment of the potential of ICT skills programmes and open data to empower women in the ICT sector in Kosovo." In *Situating Open Data: Global Trends in Local Contexts*, edited by Danny Lämmerhirt, Ana Brandusescu, Natalia Domagala, and Patrick Enaholo. Cape Town: African Minds, 2020.

EC. 2022. "EU Air Quality Standards." Available at https://environment.ec.europa.eu/topics/air/air-quality/eu-air-quality-standards_en. Accessed Aug 17, 2022.

Edwards, Rufus. "Impact of Air Pollution on Health and Measures to Reduce Exposures in Kosovo." UNICEF, 2020.

European Commission for the Efficiency of Justice (CEPEJ). "European Judicial Systems: CEPEJ Evaluation Report." 2018. Strasbourg, France: CEPEJ and Council of Europe. Available at https://rm.coe.int/evaluation-report-part-1-english/16809fc058.

Fedushko, Solomiia Stepanivna, Nataly Ortynska, Yuriy Syerov and Ruslan Kravets. "E-law and E-justice: Analysis of the Switzerland Experience." *CybHyg*, 2019.

Fu, Shihe, V. B. Viard, and Peng Zhang. "Air Pollution and Manufacturing Firm Productivity: Nationwide Estimates for China." *Munich Personal RePEc Archive*, 2017.

Global Open Data Index (GODI). "Place Overview." 2015. Available at http://index.okfn.org/place.html. Accessed July 15, 2022.

Gonzálvez-Gallego, Nicolás and Laura Nieto-Torrejón. "Can Open Data Increase Younger Generations' Trust in Democratic Institutions? A Study in the European Union." *PLOS ONE*, January 2021a, vol. 16, no. 1, pp. e0244994. https://doi.org/10.1371/journal.pone.0244994.

Gonzálvez-Gallego, Nicolás, and Laura Nieto-Torrejón. "Government Data Openness and Coverage. How Do They Affect Trust in European Countries?" *Journal of Data and Information Science*, February 2021b, no. 6, vol. 1, pp. 139–153. https://doi.org/10.2478/jdis-2021-0010.

Gonzálvez-Gallego, Nicolás, Laura Nieto-Torrejón & María Concepción Pérez-Cárceles. "Is Open Data an Enabler for Trust? Exploring the Link and the Mediating Role of Citizen Satisfaction." *International Journal of Public Administration*, vol. 43, no. 14, September 2019, pp. 1218-1227. https://doi.org/10.1080/01900692.2019.1668412.

Gurajala, Supraja, Suresh Dhaniyala, and Jeanna N. Matthews. "Understanding Public Response to Air Quality using Tweet Analysis." *Social Media + Society*, vol. 5, no. 3, 2019, pp. 2056305119867656.

Hanna, Rema, Bridget Hoffmann, Paulina Oliva, and Jake Schneider. "The Power of Perception: Limitations of Information in Reducing Air Pollution Exposure." *IDB Working Paper Series*, 2021.

IDRA. "Data Quality Review: Millennium Foundation Kosovo." June 2022.

International Management Group (IMG). "Kosovo Judicial Council - Case Management Information System." 2013. Available at https://www.img-int.org/project/kosovo-judicial-council-case-management-information-system-nor106. Accessed May 13, 2022.

Janke, Katharina. "Air Pollution, Avoidance Behaviour and Children's Respiratory Health: Evidence from England." *Journal of Health Economics*, vol. 38, 2014, pp. 23-42.

Khutkyy, Dmytro. "Electronic Democracy in Belarus, Moldova, and Ukraine. Patterns and Comparative Perspectives." *Comparative Southeast European Studies*, vol. 67, no. 2, 2019, pp. 264-284. https://doi.org/10.1515/soeu-2019-0017.

Koleros, Andrew, Sean Mulkerne, Mark Oldenbeuving, et al. "The Actor-Based Change Framework: A Pragmatic Approach to Developing Program Theory for Interventions in Complex Systems." *American Journal of Evaluation*, vol. 41, no. 1, Mar. 2020, pp. 34–53. https://doi.org/10.1177/1098214018786462.

Koleros, Andrew and John Mayne. "Using Actor-Based Theories Of Change to Conduct Robust Contribution Analysis in Complex Settings." *The Canadian Journal of Program Evaluation,* vol. 33, no. 3, 2019. https://doi.org/10.3138/cjpe.52946.

Krasniqi, Besnik. "Investors' Perceptions of Kosovo's Business Environment: Survey with Current and Potential Investors." 2013. http://dx.doi.org/10.13140/RG.2.2.24448.74247.

Kucera, Jan, and Dusan Chlapek. "Benefits and risks of open government data." *Journal of Systems Integration,* vol. 5, no. 1, 2014, pp. 30.

Lavy, Victor, Avraham Ebenstein, and Sefi Roth. "The Impact of Short Term Exposure to Ambient Air Pollution on Cognitive Performance and Human Capital Formation." *NBER Working Paper Series*, 2014.

Lee, Hsiu-An, Hsin-Hua Kung, Yuarn-Jang Lee, Jane C-J. Chao, Jai Ganesh Udayasankaran, Hueng-Chuen Fan, Kwok-Keung Ng, Yu-Kang Chang, Boonchai Kijsanayotin, Alvin B. Marcelo, Chien-Yeh Hsu. "Global Infectious Disease Surveillance and Case Tracking System for COVID-19: Development Study." *JMIR Med Inform,* vol. 8, no. 12, December 2020, e20567. https://doi.org/10.2196/20567.

Lim, Stephen S., Theo Vos, Abraham D. Flaxman, Goodarz Danaei, Kenji Shibuya, Heather Adair-Rohani, Mohammad A. AlMazroa, et al. "A Comparative Risk Assessment of Burden of Disease and Injury Attributable to 67 Risk Factors and Risk Factor Clusters in 21 Regions, 1990–2010: A Systematic Analysis for the Global Burden of Disease Study 2010." *The Lancet*, vol. 380, no. 9859, 2013, pp. 2224-2260.

Limaye, Vijay S., Kim Knowlton, Sayantan Sarkar, Partha S. Ganguly, Shyam Pingle, Priya Dutta, Sathish L. M., et al. "Development of Ahmedabad's Air Information and Response (AIR) Plan to Protect Public Health." *International Journal of Environmental Research and Public Health*, vol. 15, no. 7, 2018, pp. 1460.

Lnenicka, Martin and Anastasija Nikiforova. "Transparency-by-Design: What is the Role of Open Data Portals?" *Telematics and Informatics*, vol. 61, August 2021, 101605. https://doi.org/10.1016/j.tele.2021.101605.

Manisalidis, Ioannis, Elisavet Stavropoulou, Agathangelos Stavropoulos, and Eugenia Bezirtzoglou. "Environmental and Health Impacts of Air Pollution: A Review." *Frontiers in Public Health*, vol. 8, 2020, pp. 14.

Matasick, Craig. "Open government: How transparency and inclusiveness can reshape public trust", in *Trust and Public Policy: How Better Governance Can Help Rebuild Public Trust*. Paris: OECD Publishing, 2017. https://doi.org/10.1787/9789264268920-8-en.

Mavragani, Amaryllis and Gabriela Ochoa. "Google Trends in Infodemiology and Infoveillance: Methodology Framework." *JMIR Public Health and Surveillance*, vol. 5, no. 2, 2019, pp. e13439.

Millennium Challenge Corporation (MCC). 2020. "Evaluation Management Guidance." Available at https://www.mcc.gov/resources/doc/guidance-evaluation-management#:~:text=As%20per%20MCC's%20M%26E%20Policy,project%20objectives%20and%20theory%20of.

Millennium Challenge Corporation (MCC). 2021. Requisition No. DPE-PR-21-0195, Attachment J.4: Project Evaluability Assessment.

Millennium Foundation Kosovo (MFK). "Millennium Foundation of Kosovo Monitoring and Evaluation Plan: Version 4." September 2021.

Millennium Foundation Kosovo (MFK). "Social and Gender Integration Plan." May 2018.

Naiker, Y., R. D. Diab, M. Zunckel, and E. T. Hayes. "Introduction of Local Air Quality Management in South Africa: Overview and Challenges." *Environmental Science & Policy*, vol. 17, 2012, pp. 62-71.

Nesta. "The Open Data Challenge Series." 2022. Available at https://www.nesta.org.uk/project/open-data-challenge-series. Accessed August 19, 2022.

NIRAS and Atmoterm. "Task D1: Update of the Emission Inventory: Detailed Methodology for the Inventory of Emissions of Substances in Kosovo, the Scope of the Electronic Database and Results." 2020.

NIRAS and Atmoterm. "Supply of Project Management, Air Quality Information Management, Behavior Change, and Communication Services: Task D3: Model and Model Outputs. Modelling Methodology and Results – Update." Contract No: RFP / MFK /2019/ QCBS / No. 006. Produced for Millennium Foundation Kosovo, March 2021.

NIRAS. 2022. "Supply of Project Management, Air Quality Information Management, Behavior Change, and Communication Services: Final Report." Contract No: RFP / MFK /2019/ QCBS / No. 006. Reporting period: 30 September 2019 – 31 March 2022. Produced for Millennium Foundation Kosovo. April 27.

O'Hara, Kieron. "Transparency, Open Data and Trust in Government: Shaping the Infosphere." *Proceedings of the 4th Annual ACM Web Science Conference*, June 2012, pp. 223–232. Available at https://doi.org/10.1145/2380718.2380747.

Oltra, Christian and Roser Sala. "Communicating the Risks of Urban Air Pollution to the Public: A Study of Urban Air Pollution Information Services." Revista Internacional de Contaminación Ambiental, vol. 31, no. 4, 2015, pp. 361-375.

Open Data Barometer (ODB). "The Open Data Barometer." 2019. Available at https://opendatabarometer.org/?_year=2017&indicator=ODB. Accessed August 20, 2022.

Pepaj, Islam. "Review of Administrative Justice in the Republic of Kosovo," *Academicus International Scientific Journal*, July 2015, vol. 12, pp. 155-168.

Pherali, Tejendra, and Arif Sahar. "Learning in the Chaos: A Political Economy Analysis of Education in Afghanistan." *Research in Comparative and International Education 13*, no. 2, June 2018, pp. 239–58. https://doi.org/10.1177/1745499918781882.

Raça, Vigan, Goran Velinov, Betim Cico and Margita Kon-Popovska, "Application-based Framework for Analysis, Monitoring and Evaluation of National Open Data Portals." *International Journal of Advanced Computer Science and Applications*, vol.12, no. 11, 2021. http://dx.doi.org/10.14569/IJACSA.2021.0121104.

Raça, Vigan, Nataša Veljković, Goran Velinov, Leonid Stoimenov, Margita Kon-Popovska. "Real-Time Monitoring and Assessing Open Government Data: A Case Study of the Western Balkan Countries." in *ICT Innovations 2020. Machine Learning and Applications*, October 2020, pp.118-201. https://doi.org/10.1007/978-3-030-62098-1_16.

Ramírez, A. Susana, Steven Ramondt, Karina V. Bogart, and Raquel Perez-Zuniga. "Public Awareness of Air Pollution and Health Threats: Challenges and Opportunities for Communication Strategies to Improve Environmental Health Literacy." *Journal of Health Communication*, vol. 24, no. 1, 2019, pp. 75-83.

Ramondt, Steven and A. S. Ramírez. "Media Reporting on Air Pollution: Health Risk and Precautionary Measures in National and Regional Newspapers." *International Journal of Environmental Research and Public Health*, vol. 17, no. 18, 2020, pp. 6516.

Rashiti, Naim. "Ten years after EULEX: Key principles for future EU flagship initiatives on the rule of law." *CEPS Paper in Liberty and Security in Europe*, no. 2019-07, May 2019. Available at https://www.ceps.eu/wp-content/uploads/2019/05/LSE2019-07_Ten-years-after-EULEX.pdf.pdf.

Reiling, Dory, and Francesco Contini. 2022. "E-Justice Platforms: Challenges for Judicial Governance." *International Journal for Court Administration*, vol. 13, no. 1, April 2022, p.6. http://doi.org/10.36745/ijca.445.

Reuters. "Legal Research: Westlaw." 2022. Available at https://legal.thomsonreuters.com/en/westlaw. Accessed August 12, 2022.

Riley, Rosie, Laure d. Preux, Peter Capella, Cristian Mejia, Yuya Kajikawa, and Audrey d. Nazelle. "How do we Effectively Communicate Air Pollution to Change Public Attitudes and Behaviours? A Review." *Sustainability Science*, vol. 16, no. 6, 2021, pp. 2027-2047.

Ruijer, Erna and Albert Meijer. "Open Government Data as an Innovation Process: Lessons from a Living Lab Experiment." *Public Performance & Management Review*, vol. 43, no. 3, February 2019, pp. 613-635. https://doi.org/10.1080/15309576.2019.1568884.

Saberian, Soodeh, Anthony Heyes, and Nicholas Rivers. "Alerts Work! Air Quality Warnings and Cycling." *Resource and Energy Economics*, vol. 49, 2017, pp. 165-185

Sadiku, Lejla and Yaera Chung. "Eastern Europe and Central Asia" *in The State of Open Data: Histories and Horizons*, edited by Davies, T., Walker, S., Rubinstein, M., & Perini, F. Cape Town and Ottawa: African Minds and International Development Research Centre, 2019.

Semenza, Jan C., Daniel J. Wilson, Jeremy Parra, Brian D. Bontempo, Melissa Hart, David J. Sailor, and Linda A. George. "Public Perception and Behavior Change in Relationship to Hot Weather and Air Pollution." *Environmental Research*, vol. 107, no. 3, 2008, pp. 401-411.

Shapiro, Joseph S. and Reed Walker. "Why is Pollution from US Manufacturing Declining? The Roles of Environmental Regulation, Productivity, and Trade." *American Economic Review*, vol. 108, no. 12, 2018, pp. 3814-3854.

Stojkov, Milan, Stevan Gostojić, Goran Sladic, Marko Marković, and Branko Milosavljević. "Open Government Data in Western Balkans: Assessment and Challenges." *The 6th International Conference on Information Society and Technology*, February 2016. https://www.researchgate.net/publication/301566986_Open_Government_Data_in_Western_Balkans_Assessment_and_Challenges.

United Nations Development Programme (UNDP). "Open Data Readiness Assessment for Serbia." Washington, DC: UNDP, December 2015. Available at https://www.undp.org/serbia/publications/open-data-readiness-assesment.

United Nations Development Programme (UNDP). "Public Pulse Brief XXI." Washington, DC: UNDP, 2021. Available at https://www.undp.org/sites/g/files/zskgke326/files/migration/ks/PublicPulse-EN.pdf.

United Nations Development Programme Kosovo (UNDP Kosovo). "Public Pulse Brief XXII." Washington, DC: UNDP, 2022. Available at https://www.undp.org/sites/g/files/zskgke326/files/2022-07/PP2022EN.pdf.

United States Agency for International Development (USAID). "Case Tracking and Management Guide." Washington, DC: Bureau for Democracy, Conflict, and Humanitarian Assistance, September 2001. Available at https://pdf.usaid.gov/pdf_docs/Pnacp336.pdf.

United States Agency for International Development (USAID). "Designing and Implementing Court Automation Projects: Practical Guidance for USAID DRG Officers." Washington, DC: USAID, August 2019. Available at https://www.academia.edu/63345929/DESIGNING_AND_IMPLEMENTING_COURT_AUTOMATION_PROJECTS_Practical_Guidance_for_USAID_DRG_Officers.

United States Agency for International Development (USAID). "Thinking and Working Politically through Applied Political Economy Analysis." Washington, DC: Center of Excellence on Democracy, Human Rights and Governance, November 2018. Available at https://usaidlearninglab.org/sites/default/files/resource/files/pea_guide_final.pdf.

US Census Bureau. "The Open Data for Good Grand Challenge Awarded $260,000 in Prizes." 2021. Available at https://opportunity.census.gov/prize-challenge/. Accessed August 17, 2022.

Velicogna, Marco, Antoine Errera, and Stéphane Derlange. "e-Justice in France: the e-Barreau experience." *Utrecht Law Review*, vol 7, no.1, 2011, pp 163-187. http://doi.org/10.18352/ulr.153.

Warrener, Debbie. "Synthesis Paper 3: The Drivers of Change Approach." London: ODI, 2004. Available at https://cdn.odi.org/media/documents/3721.pdf.

Weng, Yu-Ling, Chin-Chia Liang, Chie-Chien Tseng, Shih-Yi Lee, and Gwo-Liang Yeh. "A Survey of PM2.5 Preventive Behavioral Intention and Related Factors among Community Elderly in Northern Taiwan." *Medicine*, vol. 100, no. 29, 2021, pp. e26675.

World Health Organization (WHO). "WHO Global Air Quality Guidelines. Particulate Matter ($PM_{2.5}$ and PM_{10}), Ozone, Nitrogen Dioxide, Sulfur Dioxide, and Carbon Monoxide." Geneva: World Health Organization, 2021. Available at https://apps.who.int/iris/handle/10665/345329.

Williams, Gareth. "Thinking and Working Politically on Health Systems Resilience: Learning from the Experience of Cameroon, Nepal and South Africa During COVID-19." *The Thinking and Working Politically Community of Practice,* June 2022. Available at https://twpcommunity.org/wp-content/uploads/2022/06/TWP-about-health-systems-resilience-reflection-note-final.pdf.

World Bank. "Air Pollution Management in Kosovo." Washington, DC: World Bank, 2019. http://hdl.handle.net/10986/33041.

World Bank. "World Development Indicators." Washington, DC: World Bank, 2022. Available at https://data.worldbank.org/indicator. Accessed July 20, 2022.

You, Wan-Hai, Hui-Ming Zhu, Keming Yu, and Cheng Peng. "Democracy, Financial Openness, and Global Carbon Dioxide Emissions: Heterogeneity Across Existing Emission Levels." *World Development*, vol. 66, 2015, pp. 189-207.

Zakharov, Nikita. "Does Corruption Hinder Investment? Evidence from Russian Regions." *European Journal of Political Economy*, vol. 56, January 2019, pp. 39-61. https://doi.org/10.1016/j.ejpoleco.2018.06.005.

Zogaj, Alban, Burim Hashani, Violeta Rexha, Gender and Social Inclusion Specialist, Edona Kurtolli-Alija, Adrian Prekaj, Sarah Olmstead, Stefan Osbourne, Bradley Cunningham, Jozefina Cutura. "Kosovo Constraints Analysis." November 22, 2017. Available at https://millenniumkosovo.org/wp-content/uploads/2018/11/Constraints-Analysis-Kosovo_Final.pdf.

This page has been left blank for double-sided copying.

Annex A. Differences in text and rationale for evaluation research questions

In this section we specify any changes to evaluation research question text that we propose, along with the reason for those updates.

Table A.1. Proposed changes to evaluation research questions

Original RQ#	Original research question	Proposed research question and revised RQ#				Reason for change
		Cross-cutting	PAJI	EDC	KODC	
1	Was the program implemented according to plan (in terms of quantity and quality of outputs)? Please look separately at each activity.	N/A	1. Was the **activity** implemented according to plan (in terms of quantity and quality of outputs)?			Shifts focus to individual activities
2	Did the program achieve its targeted outcomes, particularly its stated objective, in the timeframe and magnitude expected? Why or why not?	1. [No change]	2. Did the **activity** achieve its targeted outcomes, particularly its stated objective, in the timeframe and magnitude expected? Why or why not?			Shifts focus to individual activities
3	Do the results of the program justify the allocation of resources towards it? *MCC did not conduct an ex-ante cost benefit analysis of this investment, and we do not expect a cost benefit analysis to be done in order to answer this question, though we are open to such an analysis if the learning outweighs the cost. At a minimum, a qualitative answer based on the evaluator's perspective will suffice. But we welcome other proposals to answer this question.*	2. [No change]	N/A			N/A

Original RQ#	Original research question	Proposed research question and revised RQ#				Reason for change
		Cross-cutting	PAJI	EDC	KODC	
4	Is there any increase in the Government's demand for and consumption of analyses done by nongovernment entities, both analyses supported by MFK and those in general? Is there any evidence these types of analyses can factor into any policy decision-making? Did KODC result in increased engagement between government and civil society/ private sector? Did publishing energy and judicial data through relevant government websites (ERO, Judicial Institution TBD) result in increased analysis done by NGOs?	3a. Is there any increase in the Government's **use of** analyses done by non-government entities, both analyses supported by MFK and those in general? Is there any evidence these types of analyses can factor into any policy decision-making? 3b. Has engagement between government and civil society/private sector increased?	3a. Is there any increase in the Government's **use of** analyses done by non-government entities, both analyses supported by MFK and those in general? Is there any evidence these types of analyses can factor into any policy decision-making? 3b. Did publishing **judicial data** through relevant government websites (specific judicial institutions TBD) result in increased analysis done by NGOs? 3c. Did **PAJI** result in increased engagement between government and civil society/media?	3a. Is there any increase in the Government's **use of** analyses done by non-government entities, both analyses supported by MFK and those in general? Is there any evidence these types of analyses can factor into any policy decision-making? 3b. Did the **dissemination of air quality data** through government websites affect activities by NGOs/CSOs, and if so, why? 3c. Did **EDC** result in increased engagement between government and civil society/media?	3a. Is there any increase in the Government's **use of** analyses done by non-government entities, both analyses supported by MFK and those in general? Is there any evidence these types of analyses can factor into any policy decision-making? 3b. Did publishing **energy, labor force, air quality, and judicial data** through relevant government websites result in increased analysis done by NGOs? 3c. Did **KODC** result in increased engagement between government and civil society/media?	Changes language in 4a to use of analyses which can be more readily observed than demand or consumption; Modified 4b questions are now activity specific; Changes in 4c. reflect how each activity could affect engagement between government and civil society/private sector, not just through KODC
5	Is there an improvement in citizens' use of data to advocate for change?	N/A	4. How has the project contributed to citizens' use of **judicial** data in advocating for change?	4. How has the project contributed to citizens' use of **environmental** data in advocating for change?	4. How has the project contributed to citizens' use of **open** data in advocating for change?	Shifts focus to individual activities

Original RQ#	Original research question	Proposed research question and revised RQ#				Reason for change
		Cross-cutting	PAJI	EDC	KODC	
6	Does the existence of transparent, government produced air quality data, health advisories and a national outreach and behavior change campaign create enough incentive for civilians to change their behavior (e.g. take actions to reduce the negative health impacts of air pollution)? If evidence of changed behaviors exists, who is adapting, how have they adapted and why?	N/A	N/A	5. [No change]	N/A	N/A

Original RQ#	Original research question	Proposed research question and revised RQ#				Reason for change
		Cross-cutting	PAJI	EDC	KODC	
7	Does the existence of transparent, government produced air quality data reduce the adversarial relationship between civil society and the GOK? If yes, whose attitudes and behaviors are likely contributors to these reductions? Is there an increase in communication between inter-ministerial agencies, i.e. Kosovo Environmental Protection Agency (KEPA) and Public Health Institute (PHI)?	N/A	N/A	6a. Does the existence of transparent, government produced air quality data reduce the adversarial relationship between civil society and the GOK? If yes, whose attitudes and behaviors are likely contributors to these reductions? 6b. **Has inter-ministerial communication changed**, e.g., between Kosovo Environmental Protection Agency (KEPA) and National Institute of Public Health Institute (NIPH), **and if so, why and how?** **6c. Is air pollution data available on a continuously updated basis? How accurate are the air pollution forecasts provided through the NIPH portal? What percentage of time does air quality exceed given thresholds?**	N/A	Change in wording for 7b to signal a more comprehensive treatment of inter-ministerial relations, not just a measure of communication frequency; Addition of 7c supports the evaluation of 7a by examining forecast accuracy and underlying air quality conditions observed at KHMI stations

Original RQ#	Original research question	Proposed research question and revised RQ#				Reason for change
		Cross-cutting	PAJI	EDC	KODC	
8	Is there an increase in government employees who perceive government data as a public good or resource to be shared? If yes, how are government employees sharing data with the public (open data?, website?, reports?)? If no, why?	4. Is there **a change** in government employees' perceptions of government data as a public good or as a resource to be shared? If yes, how are government employees sharing data with the public (open data?, website?, reports?)? If no, why?	N/A			Shifts focus to a qualitative treatment since quantitative data alone is unlikely to be accurate or informative
9	Did EDC, KODC, and PAJI contribute in increasing trust and understanding of government's function?	N/A	5. Did **PAJI** contribute to increasing trust and understanding of the judiciary system's functions?	7. Did **EDC** contribute to increasing trust and understanding of government's function?	5. Did **KODC** contribute in increasing trust and understanding of government's function?	Shifts focus to individual activities

Note: Bolded sections denote the specific parts of a research question that we have updated.

This page has been left blank for double-sided copying.

Annex B. Evaluability assessment tables

This annex summarizes the evaluability assessment of each TAG activity. We used various data sources for this assessment, including hundreds of programmatic documents, discussions with 20 stakeholders during an in-country evaluation design mission trip, and relevant literature. Program documents included (1) M&E plans, logic models and indicator tracking tables; (2) monthly, quarterly, and annual implementer reports; (3) meeting notes and slides; (4) implementer terms of reference; (5) analytical products; (6) product guides or manuals; (7) budget documents; (8) grantee concept notes and milestone reports, and (9) the constraints analysis. Table B.1 lists the guiding questions and sub-questions for each of the five evaluability dimensions.

Table B.1. Evaluability questions by dimension

Dimension question	Dimension sub-questions
1. Is the problem clearly defined and is there sufficient evidence to support the problem diagnostic?	**1.** Is there quantitative evidence regarding constraints to and sources of economic growth? **2.** Is the problem(s) clearly defined and understood with sufficient evidence/quantitative (baseline) data available to support claims? **3.** Is there evidence to support root causes identified? **4.** Will all root causes be addressed by the proposed intervention or other, complementary intervention(s)? **5.** Is there a public good rationale and/or market failure that necessitates government intervention and funding? This is linked to sustainability – how will recurrent costs be covered in the future, how will private sector investment be triggered, etc? **6.** Is the institutional context understood, the political economy understood, and does the team clearly understand how the proposed intervention(s) link to other initiatives by the gov't and/or other partners? **7.** Is there a clear understanding on how different social and cultural dynamics (gender, poverty, race, ethnicity, etc.) may be influenced by or influence the problem identified?
2. Are the project objectives and theory of change/logic clearly defined?	**1.** Is the objective of the Project clearly stated with a link from the problem diagnostic? **2.** Is the project logic and economic model clear, plausible and based on existing evidence and literature? If there is limited evidence, is there commitment to generating evidence via an impact evaluation (linked to Section 5)? **3.** Is there a clear logic that links different projects within the Compact program that is then linked to accelerating economic growth? **4.** Is it clear which component(s) of the problem diagnostic will be addressed by the proposed intervention, which will be addressed by complementary activities and which will remain risks for the MCC intervention to achieve proposed results? **5.** Are the inputs, outputs, outcomes clearly defined and linked to the economic analysis (ERRs)? **6.** Is the timeline for expected results clear and based on evidence? **7.** Is it clear whether or not benefits are expected to be sustained beyond the life of the compact?

Dimension question	Dimension sub-questions
3. Are the risks and assumptions clearly defined with potential risk mitigation strategies?	**1.** Are the risks to achieving expected results clear, with clearly defined risk mitigation strategies? **2.** Does the ERR reflect these assumptions and risks? Has sensitivity analysis been used to select key risks and assumptions? **3.** Is it clear how risks will be monitored? **4.** Is it clear how design and implementation may be altered as information on new risks/realization of risks occurs? **5.** Does the project team make a critical assessment of the degree to which there may be blind spots or unknown unknowns in a project of this nature (e.g. how foreseeable are the potential risks that may arise in new sectors)?
4. Are project participants clearly defined and justified in terms of geographic scope and eligibility criteria?	**1.** Is the selection criteria for project participants clearly defined and based on the problem and evidence in the program logic? **2.** Is program participants' selection based on credible, quantifiable selection criteria? **3.** Are specific demographics (age, gender, poverty status) defined where necessary? **4.** Are the geographic location(s) for the Project defined and based on the problem listed above and evidence in the program logic? **5.** Will the Project design and implementation plan vary by different sub-groups and/or geographic locations based on the problem listed above and evidence in the program logic? **6.** Can the selection be replicated for the purposes of an impact evaluation (linked with Section 5)?
5. Are the metrics for measuring results for both accountability and learning clearly defined?	Are there clearly defined indicators and data sources identified for monitoring **project implementation** and **project results** (consider separately)? **1.** Are there clearly defined indicators for measuring expected performance (processes, outputs)? **2.** Are the indicators linked to the ERR? **3.** Is it clear which indicators will be disaggregated by gender, age, income as appropriate? **4.** Is there a clear understanding of the time frame for expected results of each indicator (if varies)? **5.** Is there sufficient information to set appropriate and feasible baseline and annual/quarterly targets? **6.** Are there sufficient human and financial resources in the MCA and IEs to conduct the necessary data collection/reporting during the life of the intervention? Are data collection costs known and budgeted for? **7.** Is it clear who will use the data and for what purpose(s)?

Dimension question	Dimension sub-questions
	Is the evaluation clearly defined for **maximizing learning and accountability**? **1.** Is there commitment by all key stakeholders to implement the independent evaluation? **2.** Are evaluation questions and outcomes clearly defined and prioritized? **3.** Is it clear which outcomes will be disaggregated by gender, age, income as appropriate? **4.** Is it clear who will use the evaluation results and for what purpose(s)? **5.** Is the evaluation methodology the most rigorous and feasible possible? **6.** Is it clear how an evaluation (performance or impact) will contribute to the evidence base in the sector? **7.** Are there interim/continuous evaluation results which could help inform decisions during the compact life? If so, is such an evaluation built into the evaluation plan? **8.** Do the potential benefits and learning from an evaluation of the program outweigh the costs? **9.** Are there sufficient human and financial resources in the MCC, MCA and IEs to conduct necessary data collection/reporting during the life of the evaluation? Are data collection costs known and budgeted for?

We sought to answer these questions and sub-questions for each activity, drawing out remaining needs and implications for our evaluation. Table B.2, Table B.3, and Table B.4 present synthesized findings from the evaluability exercise for PAJI, EDC, and KODC.

1. PAJI

Table B.2. PAJI evaluability assessment

Dimension	Assessment	Needs and implications
Problem definition and diagnostic	• Moderate levels of evidence (documented in Constraints Analysis) indicate that weak real and perceived rule of law constrains economic growth. • Not all root causes of weak real and perceived rule of law will be addressed by the PAJI project. • Documents justify the PAJI investments as developing a public good, but sustainability of the activity's products once transferred to the government is not guaranteed. • Design of PAJI shows a strong understanding of institutional context, political economy, and concurrent interventions, and a moderate understanding of how social and cultural dynamics interact with the root problem.	• The problem diagnostic is adequate • Evaluation will seek more information on proposed steps to help public institutions sustainably manage PAJI products after the project
Objectives and logic model	• Evidence supports the assessment that a lack of data is one of the factors contributing to negative perceptions of the rule of law, but it was not completely clear why this specific problem was chosen to be addressed from among various other factors impacting the rule of law. • While project logic is relatively straightforward, not all assumptions embedded in the logical links are adequately explored—for example: – if data are made available but are delayed, of poor quality, or reflect a poor state of the judicial system, trust in government may not increase; and – evidence backing the assumed link between increased citizen use of judicial data and economic growth is not adequate. • Documents and stakeholders indicate the project has not adhered to the timeline specified in the deliverable schedule. • Benefits of PAJI are expected to remain after the threshold, but specifics of how recurrent costs will be covered within KJC are not clear.	• Evaluation will seek more information on how assumptions in logical links were identified and tested
Assumptions and risks	• Risks are not adequately defined and preempted with mitigation. strategies, but TAG-wide documents suggest the implementer is responsible for providing risk registers and monitoring systems in their deliverables. • Documents do not critically examine possible alterations to project design nor blind spots.	• Evaluation will seek more information on preparations for risks or program alterations
Project participants and beneficiaries	• Documents make clear that the judicial system, litigants, and the public are beneficiaries of the PAJI activity. • Stakeholders indicated improvements in case tracking mechanisms may help ethnic minorities without access to e-Kosovo credentials check on their property cases.	• Definitions of participants and beneficiaries are adequate

Dimension	Assessment	Needs and implications
Monitoring and measurement	• Indicators and data sources for monitoring PAJI implementation are sparse, mostly covering progress in platform development and training. • Indicators and data sources for assessing PAJI results are varied in quality; some are vague, as in *judicial data communication improved*, and others are adequately granular and include disaggregation. • All stakeholders appear committed to the success of the independent evaluation, and the general parameters of the evaluation are adequately listed in the monitoring and evaluation plan. • Evaluation questions and outcomes are not adequately disaggregated by gender, income, and age as written.	• Evaluation will seek more information on: • how indicators were selected; • opportunities for disaggregation; and • timeline for case digitization

2. EDC

Table B.3. EDC evaluability assessment

Dimension	Current assessment	Needs and implications
Problem definition and diagnostic	• Documents indicate poor air quality leads to health issues and deters investment in Kosovo, posing an obstacle to economic growth. • Documents adequately show how the market failure of low air quality necessitates government intervention for impact and sustainability; the institutional context is well understood but there was inadequate information on how citizens found air quality information before EDC. • Strong evidence linking transparency in environmental data to economic growth was lacking in the documents. • The activity does not address the root causes (dirty coal power stations, low energy efficiency, air pollution), but it does address transparency issues from all relevant angles. • EDC documents and stakeholders have clearly identified how social and cultural dynamics interact with the problem of air pollution and the need for related information.	• Evaluation will seek more information on how citizens found air quality information before EDC
Objectives and logic model	• Project documents suggest more data on air quality could increase transparency and public trust, change citizen behaviors, and engage civil society. While reduced pollution is not strictly part of the EDC theory of change, several documents indicate advocacy and behavior change could in turn lead to reduced pollution. • While there are clear links between EDC and the other activities (particularly KODC), the project documents and stakeholders did not fully explain the expected connection between improved availability and use of air quality data and accelerated economic growth. • The EDC timeline is reasonable to deliver expected outputs, but the sustainability of activity's effects on citizen/CSO air quality data use and advocacy is not clear.	• Evaluation will seek more information on how the project envisioned air quality data would lead to improved investments and economic growth
Assumptions and risks	• EDC identified risks and mitigation strategies before implementation, including insufficient involvement of stakeholders, insufficient data for pollution inventory, lack of GSI principles in the planning process. • The implementer has applied and adapted risk mitigation strategies as necessary throughout the activity, including in response to COVID-19.	• N/a
Project participants and beneficiaries	• EDC beneficiary groups and institutions are clearly defined and are justified based on evidence, as are geographic locations for air quality measurement. The criteria used for targeting specific areas for outreach are less clear. • The implementation strategy was uniform across all groups and areas; gender and social inclusion dimensions were only considered in the outreach and behavior change subactivity.	• Evaluation will seek more information on GSI integration and outreach location selection

Dimension	Current assessment	Needs and implications
Monitoring and measurement	• Indicators, data sources, and the timeline for monitoring EDC implementation appear adequate, and stakeholders suggest the project use monitoring data to adjust programming where necessary. • Indicators, data sources, and the timeline for assessing EDC results appear somewhat inadequate, with documents suggesting the evaluation rely on a survey (subject to recall bias) and the UNDP Public Pulse survey (which has erratic changes in responses due to COVID-19). • Not all targets were clearly justified in the documents or by stakeholders.	• To substitute for unreliable data sources, the evaluation team will draw app and website traffic and qualitative data to measure outcome results.

3. KODC

Table B.4. KODC evaluability assessment

Dimension	Current assessment	Needs and implications
Problem definition and diagnostic	• Documents clearly indicate root causes: a lack of publicly available data—and lack of funds—for organizations and individuals to generate insights and advocate for improvements hampers trust in government and economic growth; documents and stakeholders show understanding of the institutional context and political economy related to data publication and use. • KODC grant manuals indicate how subprojects in each area (air quality, energy, judicial, and labor) could stimulate civil society engagement in data use and advocacy; subproject concept notes do not provide quantitative evidence of constraints they aim to resolve. • Documents provide a rationale for an investment in public goods of information availability and data-driven advocacy; however, the project assumes government, civil society, and private sector will be able to maintain those goods once TAG funding expires but does not provide robust evidence there are resources available to do so. • KODC documents detail how dimensions like gender, poverty, race, ethnicity, and geographic location interact with the data problems identified and drive some aspects of project activities.	• Evaluation will seek more information on how KODC expects subprojects will generate lasting impacts • Evaluation will seek more information on specific data-related problems grantee subprojects aim to resolve
Objectives and logic model	• The KODC objective and logic model are clear and align with the problem diagnostic and the efforts of the other TAG activities. • The anticipated contributions of specific grantee projects to sustained changes in citizen engagement, government transparency, and data-driven decision-making culture are not clear. • The anticipated route by which the success of KODC may foster economic growth is not completely clear. • The timeline for KODC DigData Challenges varies (partially due to delays in the judicial activity), as does the timeline of individual grantees' subprojects.	• Evaluation will seek more information on the final timeline and sustainability plans of each grantee
Assumptions and risks	• While the Threshold Program Grant Agreement defines overall risks and mitigation strategies, risks and assumptions are only briefly discussed in KODC grantee documentation and are not explicitly addressed with mitigation strategies. • TAG-wide sensitivity analysis and assumptions behind ERR calculations are not well documented; KODC-specific documents do not clearly indicate how risks to program implementation will be monitored and how blind spots will be assessed. • TAG M&E plans present project management tools that use a risk register, a stakeholder coordination plan, a change management plan, and a work-breakdown structure; we do not have information on the application of those tools.	• Evaluation will seek more information on risks and mitigation strategies defined at the KODC and grantee level

Dimension	Current assessment	Needs and implications
Project participants and beneficiaries	• TAG M&E plan justifies target participants for each activity based on problem diagnostic and program logic; KODC-level documents do not provide clear selection criteria for project participants ("open data movers and shakers and opinion-makers, start-ups, civil society, the private sector, academia, journalists, designers, technology innovators, and creative problem solvers") but do highlight importance of elevating women in the data space. • Grantee subprojects provide variable amounts of information on audience/participant selection and demographic information. – Some subprojects indicate they have selected sites where the impacts of open data work will be greatest (for example, a high school where students are interested in pollution). • Geographic distribution of KODC subprojects is not pre-planned or systematic and could not be replicated.	• Evaluation will seek more information on how KODC developed participant selection criteria and how grantees chose their audiences
Monitoring and measurement	• Indicators, data sources, and the timeline for monitoring KODC implementation are defined at the activity level, but early documentation from DigData grantees does not define all indicators by which they measure progress. Some grantees provide improved monitoring information in their milestone reports. – Unlike some indicators for EDC and PAJI, none of the indicators for KODC in the ITT are disaggregated by gender, age, and income; few subprojects indicated they would provide disaggregated data. – Baseline values and expected timelines for measuring indicators are available for KODC, but many yearly targets are missing. • Indicators, data sources, and the timeline for assessing KODC results are somewhat clear; some KODC outcome indicators, particularly those which are to be measured by the evaluator, are not fully defined, and do not have targets set.	• The evaluation team will define and seek to disaggregate outcomes wherever possible using primary and grantee data

This page has been left blank for double-sided copying.

Annex C. MCC comments and evaluator responses

Reviewer Institution / Role (e.g. MCC /MCC/ M&E Lead)	Page Number (please reference the number at the bottom of the page)	Comment	Evaluator Responses
MFK/ Project Lead	3	*Table II.1 - Suggest rephrasing the sentence to: "Creating an Online Data Platform (ODP) to enable public access to aggregate judicial data as well as data disaggregated by demographics and other relevant categories* ". The data will not be only disaggregated by demographic categories, there will be also disaggregation based on the case type, case status, assignment type, subject matter of dispute etc.. The available data is also based on the defined / available categories in the CMIS. See PAJI Assessment Report	Revised.
MFK/ Project Lead	3	*Table II.1* - Suggest rephrasing to "Launching a Case Tracking Mechanism (CTM) that provides individual access to case information for authorized public users". This is because the Case Tracking Mechanism is a tool where individuals will be able to access their own case information through an authentication procedure (those authenticated, either individuals or lawyers will be authorized to view their case information/status digitally and track the progress of their case online).	Revised.
MFK/ Project Lead	3	*Table II.1* - Suggest rephrasing to "Supporting Kosovo Judicial Council (KJC) communication, outreach, and publication". Since the Ministry of Justice is not the beneficiary institution for PAJI post-Threshold	Revised.

Reviewer Institution / Role (e.g. MCC /MCC/ M&E Lead)	Page Number (please reference the number at the bottom of the page)	Comment	Evaluator Responses
MFK/ Project Lead	3	*Table II.1* - Suggest rephrasing to "Disbursing up to $1.3 million in total to DigData grantees to develop data-driven solutions for government agencies and/or the public". For consistency in the amount for KODC throughout the report.	Our understanding was that the total allocation for KODC was $1.3 m, but that a portion of that would cover administrative costs related to grant announcements, application review, grantee management, and so on. We had understood that up to 1 million of the KODC allocation would be for the grantees themselves (and based on the July 2022 KODC report, ~$1,004,932 had been disbursed). Still, we have changed the value to $1.3 as requested.
MFK/ M&E Lead	5	The 22 grants are for all four challenges. According to ITT data, MFK awarded 7 grants for the labor challenge. Twenty-two (22) applications were accepted, with seven (7) grantees in DigData Labor Force selected, four (4) in DigData Air Quality, five (5) in DigData Energy, and six (6) in DigData Judicial.	Revised.
MCC/ Evaluation Lead	5	"goal of increasing business investment" : the ToC shows the goal as poverty reduction through economic growth, which I believe is how MCC general uses the term goal, whereas I think this would be a long-term outcome or something of that sort, since the objective is earlier in the logic	We have updated the text to reflect business investment as an outcome.
MCC/ Evaluation Lead	5	"As he public and private sectors use the data" (Typo)	We have fixed this typo.
MCC/ M&E Lead	5	*Theory of Change* - suggest revising to …..The theory of change (TOC) illustrates how sub-activities of the TAG project contribute along a causal pathway to the overall goal of the Threshold Program (Figure II	We have accepted and implemented that suggested revision.

Reviewer Institution / Role (e.g. MCC /MCC/ M&E Lead)	Page Number (please reference the number at the bottom of the page)	Comment	Evaluator Responses
MFK/ Project Lead	5	*KODC.....Successful grantees were slated to receive between $1,000 and $50,000 to support their activities, with the intention that any apps that were developed as part of DigData funding would be handed off to relevant GoK ministries to ensure continued functionality beyond the period of the Threshold Program*. The Grant agreements foresee granting a royalty-free license to GoK to use these products, however the intellectual property rights remain with the grantees. In some cases, where we have identified jointly with grantees that there was no capacities (being it human, technological or financial) or just good will by the government to keep the solutions and ensure their sustainability over time we have encouraged grantees to continue maintaining their solutions or look for additional funding to scale up their solutions after their grant with MFK ends.	We have revised this section based on the information provided.
MFK/ Project Lead	5	*Theory of Change* .....Suggest adding, energy and labor force data as well, since labor force data is of utmost importance when it comes to how the public perceives the governments stand and data on unemployment.	We have drawn out those two areas of KODC more clearly in the paragraph.
MCC/ Evaluation Lead	7	"Millennials and Generation Z": just want to confirm that these terms are also broadly used in Kosovo/EU countries being referenced here. Not sure it's worth defining the terms (nor is it necessary to avoid them), but wanted to make sure this isn't too much of a idiomatic expression that might lose some readers	The paper in question comes from faculty at a Spanish university, draws data from EU countries, and suggests those terms are widely comprehensible. Authors cite the Pew Research Center definition as: "those born between 1981 and 1996 are considered as Millennials and anyone born from 1997 onward is part of Generation Z". For clarity, I have rephrased the language in the EDR using the dates.
MCC/ Evaluation Lead	8	"struggled to actualize these laws " - haven't seen the verb actualize used in English. I suppose the word technically exists, but I think this means to say "update"	We were using the term actualize to mean "to make real" or "to realize". To clarify, we've adjusted the langauge to "implement", since laws aren't exactly real until implemented, we're talking more specifically about implementation.

Reviewer Institution / Role (e.g. MCC /MCC/ M&E Lead)	Page Number (please reference the number at the bottom of the page)	Comment	Evaluator Responses
MCC/ Evaluation Lead	11	The transition between the "Even when" sentence and the "Usage of" sentence (and then the following sentence reverts back to the previous idea) make this paragraph a bit stilted.	We have clarified these sentences to improve their flow and coherence.
MCC/ Evaluation Lead	11	"little rigorous evidence " : I would tend to think there is little evidence of any kind, full stop, no?	Yes, that is a fair assessment. Rigorous evidence is absent, and there is little in other forms of evidence as well.
MFK/ GSI Lead	12	*Policy Relevance:* Might be important to evaluate if PAJI improved the trust of the citizens in the justice system, as well as if the access to data in the judiciary for civil society organizations, and other users supported the analysis of data derived from the CMIS across a critical set of categories has improved the effectiveness of the justice system in Kosovo. Although, I understand that this could be difficult the late implementation could offer an avenue for this	As indicated in the responses to other comments in this file, we will examine access, trust, and efficiency to the degree possible with portal data, qualitative information from a variety of stakeholders and beneficiaries, and online interest in judicial transparency and trust in institutions. As you suggest, the later implementation of PAJI has pushed to us do decide to collect data slightly later than originally anticipated in the proposal to be able to observe behavioral changes that could take time to accrue after PAJI activities are complete. I have added a note to that effect to the II.C.3 section on PAJI policy relevance.
MFK/ GSI Lead	14	*Table III.1*. Suggest using Kosovo Judicial Council instead of GoK, the PAJI is to help improve the efficiency of the judiciary system	We agree that KJC is the specific GoK agency that is most relevant to PAJI activities. However, we want to have a broader option to understand how other GoK entities or boards may use or engage with the data. We will keep the language as is but focus on KJC where relevant.
MFK/ GSI Lead	14	*Table III.1. RQ5* : Suggestion to replace with government function with “Judiciary system function”	We have updated instances of this RQ with "the judiciary system's functions."
MFK/ Project Lead	15	Table III.1 - 3b - The Open Data Portal and the Case Tracking Mechanism will both be hosted by KJC and CTM will also be available as a column to click on e-Kosovo portal that will redirect the user to KJC ODP platform, so if necessary instead of TBD it could be stated KJC.	Revised.

Reviewer Institution / Role (e.g. MCC /MCC/ M&E Lead)	Page Number (please reference the number at the bottom of the page)	Comment	Evaluator Responses
MFK/ M&E Lead	15	I suggest to specify “media” instead of private sector	We have updated the RQ language to indicate "media" instead of the "private sector", but note that the TAG logic model refers to collaboration and partnership between GoK and the *private sector*. As such, we retain references to "private sector" throughout the evaluation assessment and the evaluation design to signal that such relationships are relevant to the evaluation scope, even if there are few private sector entities included in our primary data collection plans.
MCC/ M&E Lead	15	Table III.1 - The research questions for PAJI might be using and/or interchanging GOK, government and judiciary branch of the government. Should it be specified for since PAJI platforms will be hosted by the KJC ?	Our preference is to keep the research questions as they are written to preserve the possibility we may want to understand how PAJI activities affect broader access to and use of judicial data by different government agencies, not just KJC.
MCC/ M&E Lead	17	**Context** *: For example, the project TOC assumes that increased collaboration and communication between GoK and civil society/private sector will contribute to increased investment in Kosovo by businesses, but how, why, and whether businesses will act on noticeable changes in that collaboration and communication is unclear.* These assumptions were based on the premise that transparency in the process and timeline for cases related to businesses e.g. conflicts, litigations, or disputes could encourage investors to trust the business environment of Kosovo. While still a loose assumption, the potential link for investors to make decisions based on transparency in the efficiency of the judicial process should not be completely ignored. Interested investors will have access to the data to support decision making or rely on CSO reports for information as needed	It sounds like this assumption is specifically about anticipated PAJI-related outcomes. We've added this in as an example of an assumption that does link activities to outcomes further along causal chain.

Reviewer Institution / Role (e.g. MCC /MCC/ M&E Lead)	Page Number (please reference the number at the bottom of the page)	Comment	Evaluator Responses
MFK/ GSI Lead	18	*General:* Suggest to consider exploring if PAJI will have an effect on women's access to justice, since the PAJI will also produce statistic gender disaggregated	Our understanding is that there will be some opportunity to explore this question with existing datasets and through qualitative data collection with organizations focused on women's rights and gender issues. However, "access to justice" is a difficult concept to operationalize and assess, and may be beyond the established research scope of this evaluation.
[Not specified by MCC]	18	*Box III- (DQR) :* Conclusions drawn from the DQR on TAG indicators needs to be revisited and adjusted accordingly. Given the current timeline on key activity implementations coupled with limited extensions on implementation contracts, MCC handover and sustainability plans - the challenges previously assumed for TAG indicators could be addressed and tracked during the Threshold Closeout period through January 2023 and beyond.	We are glad to hear that there may be time to address challenges related to TAG indicators. We have noted that in the text of the DQR box, but cannot change our current assessment without additional information on how indicators will be changed. If indicators improve, we will incorporate the improved indicators in our analysis.
MFK/ Project Lead	19	*Table IIII.4 - Evaluability component 2* : In terms of sustainability, KJC as a beneficiary of hardware/software components of PAJI has the capacities to maintain further the products and expressed will in doing so, however, as noted in previous discussions, public interest on these two mechanisms, especially on the Case Tracking Mechanism might reduce over time as the products developed through PAJI activity are dependent upon the Case Management Information System (CMIS a digital tool where judges, clerks, and court and prosecution staff enter judicial cases). If the frequency of information within CMIS is not at the desirable level and if cases take longer time to resolve, automatically the CTM will be affected and not show any difference as to the case status. These are factors that are non-dependent on PAJI interventions. Is this counted as a sustainability factor once transferred to GoK than this is a factor of concern as stated in the dimension score.	This aligns with our understanding (and stakeholders' cautions during our mission trip) that PAJI-funded systems are dependent on the CMIS being regularly updated. This does factor into our concerns about the logic of PAJI. When we assess sustainability of PAJI, we will also draw on qualitative data to assess the durability of CMIS and other factors that underpin PAJI's long-term effects.

Reviewer Institution / Role (e.g. MCC /MCC/ M&E Lead)	Page Number (please reference the number at the bottom of the page)	Comment	Evaluator Responses
MCC/ M&E Lead	19	*Table IIII.4 - Evaluability component 2:* The availability and access to key performance indicators of judicial authorities is expected to improve transparency and better understanding of the judicial system. While CTM gives individuals case progress and, the ODP that draws data from the CMIS is expected to show the performance of judges and courts in the country at any given time since 2018. A key assumption for PAJI is to improve the efficiency of judges and the PEA methodology suggested for assessing this activity should provide more context here	Though our evaluability assessment for dimension 2 will not change at this stage, we are incorporating your contribution into our description of how we will apply PEA in section III.E.1.a.
MCC/ M&E Lead	19	*Table IIII.4 - Evaluability component 3*: This has been the case. As noted the lack of documents to spell out these assumptions have made it challenging to highlight the causal linkages. However, PAJI's dependence on IM has always made the need to see an operational CTM and ODP more relevant. Now that we have a demonstration of the CTM and ODP in place (as of September 15) we imagine the assumptions and risks are more evident to support the logic and objectives.	As mentioned in our evaluability assessment section on PAJI, we will ingest additional documentation from implementation of the activity as it moves forward and as it concludes to understand the assumptions and risks more fully.
MCC/ M&E Lead	20	*Table IIII.5 - Evaluability component 2:* This is embedded in the EDC design and sustainability plans. The capacity building and behavior change exercises conducted with the various beneficiary agencies and institutions are part of ensuring continued use of EDC products. For instance KEPA will continue to maintain the AQ monitoring station equipment, emissions database and other training materials with other AQ donor agencies. The education ministry will incorporate AQ school curriculum...etc. The sustainability components of each activity is integral to the "next level" (post-objective) causal link to the economic growth outcomes of the project logic	We agree that KEPA and KHMI have named responsibilities to carry this work forward and have updated the text in evaluability dimension 2 to acknowledge that fact. We still believe this dimension merits an "Adequate" scoring because ample evidence of these bodies' interest and ability to sustain and scale the work developed during the Threshold is not yet available to the evaluation team.

Reviewer Institution / Role (e.g. MCC /MCC/ M&E Lead)	Page Number (please reference the number at the bottom of the page)	Comment	Evaluator Responses
MFK/ M&E Lead	21	*Table III.6. Evaluability Component 4:* The KODC the selection criteria is in grant manuals of each challenge: Labor: https://digdata.millenniumkosovo.org/the-call/ Air: https://millenniumkosovo.org/digdata_air/grant-manual/ Energy: https://millenniumkosovo.org/wp-content/uploads/2021/01/GRANT-MANUAL.docx Judicial: https://millenniumkosovo.org/wp-content/uploads/2021/09/GRANT-MANUAL-JUDICIAL-1.pdf	Our understanding was that while the eligibility and selection criteria for grantees was clear, the targeted beneficiaries for grantees' proposed work was not specified at the grant manual stage. We have adjusted the evaluability assessment table and narrative to recognize that grantees as participants have clear eligibility and selection criteria.

Reviewer Institution / Role (e.g. MCC /MCC/ M&E Lead)	Page Number (please reference the number at the bottom of the page)	Comment	Evaluator Responses
MCC/ M&E Lead	21	*Table III.6. Evaluability Component 4*: Missing indicators can be addressed via the M&E plan reviews. However, in terms of grantee level outcomes, these seem like that should be addressed through the evaluation in addition to the milestone reports grantees have shared. In addition, these can be crosschecked against Dig data grantee tracking documents	Dimension 4 deals with participants and beneficiaries, not outcome measurement. We assume this comment should instead refer to dimension 5, which covers monitoring implementation and assessing project results. As you suggest, we will use all M&E plans and reporting documents across KODC grantees to understand output and outcome achievement. At this point, we have not begun analysis of all grantee documents, including the milestone reports. Documents in our initial evaluability assessment review led us to believe that indicators were not set for most grantees that would allow for aggregation of those indicators in the ITT. There were also missing targets for several KODC-level indicators in the ITT. We are also concerned about the level of consistent detail available in grantee reports in terms of the gender, age, income, and ethnic group of beneficiaries of the DigData Challenge innovations. Again, we will collect, analyze, and synthesize as much information as we can, but may be limited in drawing painting a KODC-wide impact picture by the differences between grantees in terms of their activities, targeting, and disaggregated data. We have adjusted the text of the table and corresponding areas of the narrative to reflect that grantees provide more details in their milestone reports than we previously had known.

Reviewer Institution / Role (e.g. MCC /MCC/ M&E Lead)	Page Number (please reference the number at the bottom of the page)	Comment	Evaluator Responses
MFK/ M&E Lead	25	*Correlation Analysis*: The latest forecast that MFK has (as per the Handover documents) is for May 2022. Additionally, to my knowledge, forecasting as a service will stop after Niras's contract ends. This means that if KHMI does not procure services for forecasting, there will not be available data in the future for the correlation analysis.	We have added a footnote indicating that this analysis would be abridged if forecasts are no longer produced. From our visit in June, KHMI informed us that they would be carrying out the forecasts indefinitely, but we will have to pivot if this turns out not to be the case.
MCC/ GSI Lead	25	Data collection: under proposed methods it is noted the researchers will try to disaggregate data where possible but this may not be possible where data is deidentified. Another approach to get at some of these issues might be to do individual interview of FGDs with stakeholders during which questions could be asked specifically about, for example, whether vulnerable groups or minority groups have been able to meaningfully use the platforms and different data. A GSI focus should be more comprehensively integrated into the qualitative section, especially given the limitation on the quantitative data.	We have included a new paragraph under III.E.1.C that describes how we will obtain a GSI focus through our KIIs and FGDs since the ODP data cannot support disaggregated analyses.
MFK/ M&E Lead	25	*Descriptive trends analysis*: Apart from app downloads, would it be possible to elaborate or list the time series data that you intend to use descriptive trends analysis?	An elaborated list of time series data on which we will conduct descriptive trends analysis is available in the evaluation-specific design overview tables in Section III.E.
MFK/ Project Lead	26	*Table III.8:* CMIS cannot be used a source of data, CMIS is an internal case management information system, available only to judges and prosecutors to enter and save judicial case information, ODP would be the portal drawing the anonymized datasets from CMIS information, my suggestion delete CMIS as it is not a tool to be used as a quantitative data source and stick to ODP/CTM only.	We have deleted the reference to CMIS from this table.
MFK/ M&E Lead	26	*Google Analytics data*: Suggest adding the ODP as well	We have added ODP as one of the sources for collecting Google Analytics data.
MFK/ M&E Lead	26	*Table III.8:* Suggest using google analytics also for PAJI. It would be useful to track the website hits for ODP and users for CMT.	We had that reflected in the text and the PAJI-specific evaluation section, but had not added it to this table. Revised.

Reviewer Institution / Role (e.g. MCC /MCC/ M&E Lead)	Page Number (please reference the number at the bottom of the page)	Comment	Evaluator Responses
MFK/ M&E Lead	27	*…open datasets that have been published on government server*s : It would be useful to specify which government stakeholders you intend to monitor.	The example in the text is KPC, and we have added in that we plan to also work with contacts at KHMI, KEPA, and NIPH, among other agencies, to identify new open datasets that have download information that could be useful to understand public engagement with government data.
MFK/ GSI Lead	27	*Limitations on data disaggregation by gender and ethnicity:* This is a potential issue., however instead of that, the evaluation team could assess whether the data set was disaggregated by gender or other disaggregation from the open data set and whether their reports from the research resulted in questions related to women's or minorities issues	We can do that to the degree possible and have added language to that effect.
MCC/ Evaluation Lead	29	Similar to above for social media content analysis	We agree that a hypothesis-driven approach is also appropriate here when using social media data, but note that our treatment of this data will be exclusively descriptive with no attempt at making causal statements. We are interested in observing whether interest/usage levels are rising/falling over time which will be underpinned by our qualitative data collection.
MCC/ M&E Lead	29	Context: *…deidentified data contained in the OPD…..... Based on conversations with stakeholders, we believe the deidentified data will be shared with the evaluation team, so that we can examine individual case records rather than losing information in an aggregation process*. Given the personalized nature of individual cases and cyber-security concerns, it is expected that all data coming from the CTM will be at the aggregate level and should be sufficient for the purpose of the evaluation. Citizens are can only access their data using a government assigned "e-Kosovo ID" so specific details for most cases might not be available for the evaluation.	That is unfortunate. Conversations with stakeholders had led us to believe it would be possible to use deidentified data for the PAJI analysis that would at once protect individuals' information and allow for more granular analysis. But we understand concerns with cybersecurity. We have revised to indicate we plan on using aggregate data from all PAJI systems.

Reviewer Institution / Role (e.g. MCC /MCC/ M&E Lead)	Page Number (please reference the number at the bottom of the page)	Comment	Evaluator Responses
MFK/ GSI Lead	30	*Table III.9. (RQ2 Indicators):* Suggest to consider "improved citizens access to justice", I don't think the data will improve decision making of the citizens but access to information can improve in general access to justice to citizens.	We have separated out the two items as you suggested: • Improved access to justice for citizens • Data used for decision making by government and other relevant actors
MFK/ M&E Lead	32	*Table III.9. (RQ3a data source):* To my understanding, google trends looks at issues that Kosovo citizens are searching (demand side), whereas RQ3a, RQ3b, and RQ3c are asking about Government's usage of what was produced from NGOs. My suggestion would be to attempt to answer these RQs through KIIs and FGDs with government stakeholders only.	While RQ3 a, b, and c will be explored in large part by qualitative data and methods, we believe that we may still be able to gather some insights from secondary quantitative data, such as portal usage activity and online discourse (captured through google trends). RQ3 b does not just ask about government's use of NGO products--it asks specifically about NGO's use of published data to conduct analysis. This is something that could be informed by capturing trends in searches and downloads of datasets.
MFK/ M&E Lead	32	*Table III.9. (RQ3a data source)*: Suggest using ODP and CMT website data, focusing on (Excel and PDF) reports downloaded from both platforms.	We have added this data source for RQ3.
MFK/ GSI Lead	32	*Table III.9. (RQ4 Indicator):* Suggest to include if civils society organization is using disaggregation data related to gender, ethnicity or age or other available disaggregated data in doing analyses to identify shortcomings of these groups in their access to justice	We have added as an indicator "CSOs' use of disaggregated data to identify and remedy cross-group differences in access to justice" for RQ4.
MFK/ M&E Lead	32	*Table III.9. (RQ2 Indicators):* Suggest adding "increased efficiency in the judicial system"	We have included "Changes in judicial efficiency" as an indicator for RQ2.
MFK/ GSI Lead	32	Table III.9. (RQ2 Indicators): consider replacing government actors with judiciary or KJC. Suggestion for the PAJI to replace word government with judiciary and/or judicial institution, or KJC	We have replaced "government actors" with "judiciary."

Reviewer Institution / Role (e.g. MCC /MCC/ M&E Lead)	Page Number (please reference the number at the bottom of the page)	Comment	Evaluator Responses
MFK/ M&E Lead	32	*Table III.9. (RQ3a Indicator): Improved relationship between GOK and CS due to increased data access and government transparency* For PAJI, increased data access and government transparency sometimes may emphasize inefficiencies in the judicial system . When data will become public through ODP, the judicial system may face a lot of criticisms from NGOs that will use the data. So, when creating the instruments, I would suggest to also take into account that the Gov-NGO relationship may be adversarial in the beginning.	We acknowledge that transparency may reveal inefficiencies and thereby support critiques of judicial systems. However, we plan to keep our qualitative data collection instruments neutral in language so as to avoid priming the interviewees and focus group participants. If transparency has increased or decreased adversarial aspects of Gov-NGO relationships, we will likely capture that.
MCC/ M&E Lead	32	Table III.9. (RQ2 data sources): The MCC Scorecard could be used as a baseline reference as of 2019. Since Kosovo was not candidate for compact eligibility, this data has not been updated. However, indicators of relevance such as rule of law under the "Ruling Justly" category could be sourced from World Bank/ Freedom House and Brookings	We had already addressed this issue in our table note about the scorecards: "Since MCC Kosovo Scorecards were maintained only until to 2019, we will use the scorecards' sources for 2020, 2021, and 2022: IMF WEO (for Inflation, Fiscal Policy), Freedom House/CLD (Political Rights, Civil Liberties, Freedom of Information), World Bank/Brookings WGI (Control of Corruption, Regulatory Quality, Government Effectiveness, Rule of Law, Gender in the Economy), IFAD/IFC (Land Rights and Access, Access to Credit, Business Start-Up), Heritage Foundation (Trade Policy), WHO/UNICEF (Health Expenditures, Immunization Rates), UNESCO (Primary Education Expenditures, Girls' Secondary Education Enrollment Rate), CIESIN/YCELP (Natural Resource Protection, Child Health)."

Reviewer Institution / Role (e.g. MCC /MCC/ M&E Lead)	Page Number (please reference the number at the bottom of the page)	Comment	Evaluator Responses
MFK/ M&E Lead	32	*Table III.9. (RQ4 Indicator)*: Apart from usage, I would suggest to also explore the availability of data. The availability of data is a precondition for its use to advocate for change. Therefore, I would suggest for Mathematica to first review and assess the data that will be made public through ODP and the data that will be displayed for individual users for CMT. Once there is a comprehensive overview of the extent to which data has been made public, then Mathematica can continue to assess its usage.	As you suggest, we are planning to assess data publication and availability through implementation analysis. With the available evaluation budget, we may not be able to build out a retrospective record of data access in the past, but can build a snapshot of data availability at this point. Using that, as you suggest, we can contextualize our findings on data usage.
MFK/ GSI Lead	33	*General:* Please ensure that focus groups are diverse in terms of gender, as well as the participation of individuals from the minority groups.	We'll seek a broad recruit a cross-section of Kosovan society for participants in our focus groups. This will include, as you suggest, adequate representation of people across genders and minority groups.
MFK/ Project Lead	34	*Sample Study*: Suggest adding "and lawyers" since CTM does allow lawyer access to case information. I believe that besides individual access of private citizens engaged in a court case, a more specific search of the case would be done by the lawyer/representative of the party in the case.	We agree that since legal representatives will be able to access CTM data they should be considered in the evaluation; we will be running focus groups in which prosecutors will participate. Given that PAJI aims to transform the public's views of the judiciary, we think the sampling frame should prioritize parties to the case and not their lawyers.
MCC/ M&E Lead	34	*Study Sample:* This will be relevant to capture the gap between design and implementation. Current implementation has involved about 10 rule of law CSOs in the initial assessment exercise for the ODP development. These CSOs and media outlets are most likely to be the beneficiaries of any capacity building or trainings outside of the key judicial process actors. The targeting sample could include other NGOs for potential spillovers or outreach effects	We will include CSOs that were not involved in the PAJI design/implementation to assess for spillover effects and have clarified this in Table III.10.
MFK/ M&E Lead	35	*Table III.10 - Potential Participants*: Suggest adding the Norwegian Embassy to the other donors, since they have been the donor organization supporting CMIS	We have added the Norwegian Embassy to our list of prospective interviewees.

Reviewer Institution / Role (e.g. MCC /MCC/ M&E Lead)	Page Number (please reference the number at the bottom of the page)	Comment	Evaluator Responses
MFK/ M&E Lead	35	large--may not be able to do for budget reasons. Are we not already covering some of these groups in interviews? What, also, do these extra people really add for us in terms of information?	I have inserted the following description of our recruitment process: "We plan to recruit three to five judges and clerks from Pristina Basic Court, the Court of Appeal, and Peja Basic Court to gather insights on the ease of use of the CMIS and CTM platforms and the perceived relationship (and level of trust) between citizens and the judiciary. To recruit these members of the judiciary, we will work with our consultant Dr. Ariana Qosaj-Mustafa to identify judges in these courts through the ODP, and will then ask willing judges to suggest one to two clerks with whom we can request an interview as well. These clerks could serve the same judges we will interview or may serve other judges, but our priority will be identifying clerks with longer tenure to understand their experiences across multiple data management systems."

Reviewer Institution / Role (e.g. MCC /MCC/ M&E Lead)	Page Number (please reference the number at the bottom of the page)	Comment	Evaluator Responses
MFK/ M&E Lead	36	*FGDs:* Suggest adding FGDs with: 1. Judges and Clerks that input data on CMIS – to learn about their experiences with CMIS and any interdependencies that might affect ODP and CMT. 2. Citizens that have cases and have created accounts in CMT – to learn about their experience with using CMT 3. Lawyers that have accounts in CMT and track their clients' cases through CMT – to learn about their experiences with CMT 4. Prosecutors – 5. CSOs and Media working in the judicial area	1. We are covering judges and clerks in interviews, and anticipate those will provide a rich enough view of their experiences. 2. With our limited budget, we may not be able to add more FGDs with citizens without sacrificing other interviews and focus groups. If other interviews or FGDs do not appear essential, let us know and we will consider substituting them. We hope our FGDs with CSOs and interviews with media will provide insights into the citizen experience. 3. With our limited budget, we may not be able to add more FGDs with lawyers without sacrificing other interviews and focus groupss. If other interviews or FGDs do not appear essential, let us know and we will consider substituting them. However, we conducting focus groups with prosecutors, which may help us understand how lawyers use CTM. 4. We are conducting FGDs with prosecutors already. 5. We are conducting FGDs with CSOs and interviews with media in the judicial space already.
MFK/ M&E Lead	39	*Analysis Plan*: Moonshot only focused on KODC, so I do not think that their interviews would be useful for PAJI.	We have removed the reference to Moonshot interviews.

Reviewer Institution / Role (e.g. MCC /MCC/ M&E Lead)	Page Number (please reference the number at the bottom of the page)	Comment	Evaluator Responses
MCC/ M&E Lead	39	*Timeline and Exposure:* PAJI is expected to be launched mid-October 2022. The portal is currently being tested for the launch date. There is a concern of how much time should be considered for the trail phase but that should be fleshed out before the launch. In addition, contracts for PAJI contractors will be extended beyond September 30, so the timeline and exposure period for PAJI might shift slightly. Can the evaluation team suggest the ideal exposure period necessary robust CAPT ?	Conceptually speaking, the ideal exposure period is one that is long enough for users to have adjusted their behavior, but not so long that other confounding factors may be responsible for such changes. There is not a literature on what that duration is because the factors are very specific to the intervention and to how affected stakeholders respond (if they respond). We believe it safe to assume that it will take time for those affected by the judicial system to update their opinions about the judiciary and their experience with the courts. While the individuals directed involved in cases will be the people first to update their beliefs, they represent a small minority of the entire country - it will require more time for the citizenry writ large to update their beliefs based on other people's experiences. Since there is a further delay of training up all users of the system in its operations, our approach will be to postpone PAJI primary data collection until as late as possible in order for system-wide changes to have a greater, more perceptible effect.
MFK/ M&E Lead	41	*Table III.12. (RQ3a and 3b Indicators)*: Similar to the comment for PAJI, data transparency may not improve the relationship between gov. and civil society. In some cases, data transparency may highlight existing problems, such as air pollution, which may worsen the relationship between GoK and NGOs.	We have updated language for the indicators to make it clearer that changes in relationships could be negative or positive.
MFK/ M&E Lead	41	Table III.12. (RQ4 data source): Suggest including both online portals, the KHMI as well as the NIPH. Also, I would suggest including data from the mobile app “AQ in Kosovo”	We have updated the data sources for RQ4 to include those.

Reviewer Institution / Role (e.g. MCC /MCC/ M&E Lead)	Page Number (please reference the number at the bottom of the page)	Comment	Evaluator Responses
MFK/ M&E Lead	42	*Table III.12. (RQ5 data source)*: To assess behaviors of citizens (if and how they changed behaviors based on air quality information), I would suggest conducting FGDs with various target groups.	We have added text in III.E.2.C clarifying that we will include individuals from targeted groups in our focus groups.
MFK/ M&E Lead	42	*Table III.12. (RQ2 data source):* Suggest adding "Data from online portals created by TAG Project (usage statistics – breakdown by mobile, by desktop computer, timing of when they're looking – when pollution levels are particularly high)"	We have added those details to Table III.12.
MCC/ M&E Lead	42	*Table III.12. (RQ5 Indicators): Awareness of potential environmental health threats. Adoption of clean fuels and technologies for cooking, light, and heating* These indicators could potentially rely on data from the Energy Project component of the Threshold. MFK outreach campaign and activities on behavior change towards cleaner energy and efficiency cross cuts TAG as well so information on health awareness and uptake of clean fuels and technologies might be available on other MFK reports. Will this be relevant for the TAG evaluation ?	This appears to be a duplicate comment.
MCC/ M&E Lead	42	Table III.12. (RQ5 Indicators): Awareness of potential environmental health threats. Adoption of clean fuels and technologies for cooking, light, and heating These indicators could potentially rely on data from the Energy Project component of the Threshold. MFK outreach campaign and activities on behavior change towards cleaner energy and efficiency cross cuts TAG as well so information on health awareness and uptake of clean fuels and technologies might be available on other MFK reports. Will this be relevant for the TAG evaluation ?	We agree that data collected as part of RELP are conceptually pertinent to the TAG evaluation and should be used where possible, but from our reading of the ITT, the M&E plan, and IMPAQ's EDR it appears that RELP data will only be collected for areas directly benefiting from RELP activities. Observed changes over time would be due to the combination of RELP and TAG impacts, with no clear way to disentangle each's relative contribution. Therefore this data would not be able to provide us information on the effects of TAG alone. We would not be able to extrapolate from this sample to make any claims about whether other parts of the country that did not participate in RELP have or would have similar outcomes as indicated in the RELP data.

Reviewer Institution / Role (e.g. MCC /MCC/ M&E Lead)	Page Number (please reference the number at the bottom of the page)	Comment	Evaluator Responses
MFK/ M&E Lead	47	*Table III.13. (Stakeholder type):* Suggest including an interview with Mendim Rugova, the host of the weather forecast at Klan Kosovar. Klan Kosova was the first Kosovar TV channel to include air quality reporting based on the EDC data.	We are planning to interview Mendim Rugova whose "Mendimi Për Motin" show is now broadcast on T7.
MFK/ M&E Lead	47	*Table III.13. (Stakeholder type*): Suggest including FGDs with people who use the app and the portals to answer RQ5.	That is right - we will be recruiting individuals who have used the AQ data sources to form part of the FGDs, which will help us answer RQ5. We have refined the first illustrative theme addressed in Table III.13 for the FGDs.
MCC/ M&E Lead	52	Table III.15. (RQ3a,3b,3c data sources): Suggest listing the specific indicators of relevance from the UNDP Pulse data	PPS may serve as a data source for multiple questions across multiple TAG activity evaluations. For this KODC RQ and others, we are considering exploring indicators from PPS related to • Satisfaction with executive, legislative, judiciary • Who is most responsible for Kosovo's political situation • Satisfaction with current economic situation • Biggest problem faced by Kosovo
MCC/ M&E Lead	53	Methodology: It is not immediately clear whether the proposed 4 case studies represents one grantee from each data challenge? Given the diversity in the grantees it might be useful to increase the number of case studies to cover the different themes and milestones delivered through KODC	We have specified that indeed we are planning to conduct one case study with each DigData Challenge. To properly conduct case studies of more than 4 grantees would require more time and resources than are available. However, we have detailed our plans to capture the diversity of grantees' experience by conducting FGDs with all grantees (see Focus Groups in Table III.16).
MFK/ M&E Lead	54	*Study Sample :* "For the evaluation, we will primarily focus qualitative data collection on individuals and organizations who have directly engaged with tools supported through the Threshold Program." This is not included in Table III.16 where all target groups are specified.	We have adjusted the language in that paragraph to signal that we will collect qualitative data from those who directly engaged and those who did not directly engage in the Threshold Program.

Reviewer Institution / Role (e.g. MCC /MCC/ M&E Lead)	Page Number (please reference the number at the bottom of the page)	Comment	Evaluator Responses
MFK/ M&E Lead	56	*Table III.16. (Potential participants):* Suggest adding media outlets that write and publish articles online. Since most of the analytical news are in a written format, I would suggest interviewing them instead of the TV channels mentioned here.	We have noted our intention to engage those outlets in the table. We will work with our in-country consultants and remaining MFK staff to identify these written media sources.
MFK/ M&E Lead	61	*Table III.18. (RQ2 data source)*: Please contact MFK Leads to also send you the financial data for each TAG activity, so that you are able to obtain the resources spent.	We were planning to do this soon as part of our early data and document collection process, and I have now listed this information source in the table.
MCC/ Evaluation Lead	26-27	Seems like it would be important to have some degree of hypothesis which is being tested by the Google Trends data. While not a perfect tool, some degree of registration of the research team's priors seem important to avoid simply looking at the data for confirmation of impact. I.e.. if you find a big spike like in the shown sample data and say "oh, look, that must have been when X happened" to match what we see in the data vs. "we would expect there to be changes at X and Y dates, now lets look at the data and see whether it matches expectations". Even if it doesn't match expectation but follows a different pattern, the analysis is helped by inclusion of hypotheses	We agree that it is crucial to not engage in fishing expeditions when interpreting the data. In the activity-specific evaluations we unpack more thoroughly how we plan to approach working with this data. As you mentioned, we have key milestones that we ex ante imagine will be influential in search behavior (e.g., major launches of a new data product, press releases for a specific event). If those dates/events are not correlated with above-average search intensity, we will consult our interviewees for alternative hypotheses for why spikes occurred at specific time periods.

Reviewer Institution / Role (e.g. MCC /MCC/ M&E Lead)	Page Number (please reference the number at the bottom of the page)	Comment	Evaluator Responses
MFK/ Project Lead	3,4	*Consider updating with this revision* - .....PAJI consists of three sub-activities: (1) creating an Online Data Platform (ODP) for the public to access judicial data disaggregated by different demographic and other indicators; (2) developing a Case Tracking Mechanism (CTM) to allow individual access to case information for authorized public users; and (3) supporting communications between judicial and legal institutions and to the public. Both ODP and CTM will draw data from the Case Management Information System (CMIS), which was developed with support from the Norwegian government to increase judicial efficiency by helping judges organize their extensive caseloads. The ODP and CTM aim to enable public access to judicial data, as well as help citizens access their individual case information online to increase judicial transparency. Kosovo Legal Services Company (KLSC) in Consortium with B&S Europe were contracted to deliver the assessment and supervision of PAJI activity, whereas InfoSoft Systems sh.p.k, (in joint venture with Edusoft d.o.o., Nextsense Ltd., and Infosoft Systems sh.p.k. Albania) were contracted to implement hardware and software components of these sub-activities. The hardware and software implementation of PAJI activities was supposed to take place over a one year period from 2021 to 2022 before the Kosovo Threshold Program end date of September 30, 2022. Following the end of the Threshold Program, KJC will be responsible for maintaining the CTM and ODP along with the trainings and engagement with public stakeholders	Revised.
MCC/ GSI Lead	34 and 37	Tables with potential interviewees for PAJI and data sources: I don't see among the listed stakeholders any stakeholder working on women's justice, GBR or land issues which are key themes of interest, or any themes on questions of differential access to justice, and access and responsiveness to the needs of vulnerable groups, minorities and women. The CSOs targeted for data collection should include groups that focus on rule of law issues pertaining to women and minorities groups.	We do have these stakeholders listed in the "focus groups" section of *Table III.10. List of potential interviewees and FGD participants for PAJI evaluation.* These groups include Kosovo Women's Network, Center for Gender Studies, Roma, Ashkali, and Egyptian Women Network, Youth Initiative for Human Rights, Aktiv, Kosovo Women's Chamber of Commerce – G7

Reviewer Institution / Role (e.g. MCC /MCC/ M&E Lead)	Page Number (please reference the number at the bottom of the page)	Comment	Evaluator Responses
MCC/ GSI Lead	4, 5, 6, 7	Project descriptions as well as the lit review that follows it should incorporate a bit more focus on GSI areas of interest, e.g. that the labor force data specifically encouraged women's labor force issues analysis; that air quality component emphasized vulnerable groups and carried out dedicated messaging; that judicial data emphasizes analysis by key disaggregated variables of interest to GSI. You can pull this from the SGIP or final GSI summary report on MFK's site.	We have reviewed the SGIP and added in these important points to the project descriptions. We had included the literature we can locate on these topics (such as youth versus adult trust in institutions), but little high-quality evidence exists for some of these intervention types in terms of GSI issues.
MCC/ GSI Lead	40, 41	EDC Design: it's also missing the GSI focus. At a minimum, differential impacts of environmental data information and possible differences in citizens' usage based on for example age, which was the focus of outreach and capacity building should be integrated.	We have added information about our recruitment approach for the focus groups to ensure representation from targeted groups. We also mentioned in Table III.13 that understanding the experiences of underserved/vulnerable populations in engaging EDC outputs will be a key theme for our interviews with CSOs/NGOs. Since the quantitative data is anonymized (i.e., we don't know the sex of the person accessing an EDC portal or AQ app), we won't be able to present disaggregated statistics for the quantitative indicators.
MCC/ Evaluation Lead	Annex A	Looks great, love the proposed edits to the original research questions and the clear justifications/explanations for the same	Thank you for this feedback.

Reviewer Institution / Role (e.g. MCC /MCC/ M&E Lead)	Page Number (please reference the number at the bottom of the page)	Comment	Evaluator Responses
MCC/ Evaluation Lead	Chapter 3	Overall comment: I think the design would benefit from a careful, but creative review of the methods used by data intermediary enterprises in the private sector, how they judge user-friendliness, how they judge use of systems vs. noise in user/traffic data, etc. I may have missed it, but for example, if you think of what a research team would be tasked with if it were hired by the actual owners of these systems or owners of the data, it would likely look quite different. For example, a hands-on testing/demo testing/beta testing group of systems/platforms by actual users or potential users with observation/surveying led by the research team could be far more useful than a traditional focus group in which users simply discuss their experiences.	In our implementation analysis and thematic analysis and triangulation, we will be examining the quality of data interfaces and applications through the experiences of participants and stakeholders. These experiences and perceptions can tell us about user-friendliness aspects of specific systems, and interviews with app developers and administrators can help us understand traffic data. Unfortunately, setting up testing systems for all applications and platforms developed and used under TAG is beyond the resources for this evaluation allow.
MCC/ GSI Lead	General	The research questions do not include any that would allow us to tease out differential impacts or usage by certain groups. it would be really helpful to think of a few questions to add on that front.	The research questions are based on the original scope of work laid out for this evaluation. We hope to capture differential impacts or usage by certain groups in our current approach, particularly through our qualitative methods, but adding additional research questions would require a re-assessment (and likely addition) of methods and data sources, which would incur additional costs.
MCC/ M&E Lead	General	Formatting - minor formatting issues with figure and text alignment	We have worked with our production team to check minor formatting issues with figure and text alignment.

This page has been left blank for double-sided copying.

Mathematica Inc.

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Woodlawn, MD • Washington, DC

mathematica.org website

EDI Global, a Mathematica Company

Operating in Tanzania, Uganda, Kenya, Mozambique, and the United Kingdom