

Find the Needles, Analyse the Haystack

Lessons in Journalist-Researcher
Collaborations from Phase One
of the Illicit Finance Data Lab



Organized Crime
and Corruption
Reporting Project



Find the Needles, Analyse the Haystack: Lessons
in Journalist-Researcher Collaborations from Phase
One of the Illicit Finance Data Lab

March 2026

© Anti-Corruption Data Collective, Organized Crime
and Reporting Project

CC BY-SA 4.0

<https://creativecommons.org/licenses/by-sa/4.0/>

Authors: Alexandra Gillies, Michael Hornsby,
Hugo Williams

Research: Michelle Kendler-Kretsch,
Mara Wesseling

Editor: Lauren Jackman

Design: Joanna Poupaki / [Spoovio](#)

The Illicit Finance Data Lab team would like to thank
all the workshop participants and interview partners
for sharing their time and expertise with us.

To share any comments, feedback or
requests regarding this report, please contact
the Anti-Corruption Data Collective on
team@acdatacollective.org or the Organized Crime
and Corruption Reporting Project on info@occrp.org.

The Anti-Corruption Data Collective (ACDC)
leverages public and private data to expose
transnational corruption, illicit financial activity and
corporate opacity. Our research and investigations
seek to document and reduce the harms that
corruption causes to human security, democratic
participation and environmental sustainability: to
people, politics and planet.

To do this, ACDC brings together leading
investigative journalists, academics, policy experts
and data scientists for collaborative projects that
cross borders and break down organizational silos.
ACDC is a fiscally sponsored project of the Fund
for Constitutional Government, a 501(c)(3) nonprofit
organisation registered in Washington, DC.

**The Organized Crime and Corruption Reporting
Project (OCCRP)** is one of the largest investigative
journalism organizations in the world, headquartered
in Amsterdam and with staff across six continents.
We are a mission-driven nonprofit newsroom that
partners with other media outlets to publish stories
that lead to real-world action. At the same time, our
media development arm helps investigative outlets
around the world succeed and serve the public.



This project was funded by UK
International Development

Contents

Executive Summary	2
Introduction	4
I. The Importance of Collaboration in Tackling Financial Crime and Corruption	6
II. The Data Lab Response	8
Forging a Community	
Advancing Collaborative Projects	
Filling Knowledge & Capacity Gaps	
III. Navigating Challenges to Collaboration	13
Ethical Frameworks	
Legal Risks	
Differing Timelines	
Research Approaches	
BOX: Examples of Effective Collaboration	20
IV. Data: Navigating a Complex Landscape	22
V. Leading Themes	24
Cryptocurrency	
Beneficial Ownership and Property Data	
Professional Enablers	
Conclusion	28

Executive Summary

Financial crime and corruption drain trillions of dollars annually from economies worldwide, undermining public trust and starving essential services of funding.

G7 nations have [acknowledged](#) the scale of the problem, which they say “threatens the stability and security of societies, undermining the institutions and values of democracy, ethical values and justice”.

Today, thanks to the increasing availability of massive public and leaked datasets, we have an unprecedented opportunity to devise smarter approaches to fighting organised crime and corruption.

Combined with dramatic technological advances in research methods, these data can lead to new ways to detect, deter and ultimately dismantle kleptocratic and criminal networks and the infrastructure they depend on.

But these breakthroughs won't happen without new forms of coordinated action.

To avoid being overwhelmed by this abundance of data, journalists and academics who specialise in dirty money investigations will need to work together as never before, combining their different skills and knowledge to unlock new findings.

The Illicit Finance Data Lab was established to help scale up meaningful collaboration between these two groups, which until now has been relatively rare. If these two groups of actors can use leaked and public data to produce stronger reporting and research, then policymakers, law enforcement and the private sector will have the evidence and insight they need to act more effectively.

Led by the Organized Crime and Corruption Reporting Project (OCCRP) and the Anti-Corruption Data Collective (ACDC), the Data Lab creates the conditions for greater collaboration in data-led investigations in this emerging field by:

- building a bridge between journalists and researchers, so they can share strategies and find others doing work that is complementary to their own;
- pursuing actual investigative and research projects by raising awareness of available data and forging cross-stakeholder teams;
- filling knowledge gaps with guidance, training and lessons learned, developed with extensive input from the community.

The quick uptake of the Data Lab model suggests high demand for this type of collaboration. In just three months since our launch in January 2026, we have:

- advanced 11 investigative collaborations as a direct result of data and connections provided through the Data Lab;
- assembled an inventory of public and leaked datasets that could form the basis of collaborative projects;
- organised a two-day workshop in Amsterdam that brought together 37 anti-corruption experts from across journalism and academia to examine public and leaked datasets on priority themes, and spark future collaborations;
- reached new research and journalist communities via interviews, three online seminars co-hosted by the Governance & Integrity Anti-Corruption Evidence (GI-ACE) research programme and the University of Salford's Counter Evidentiary Network, and conference presentations;
- launched the Data Lab's members-only web platform where we published guidance articles, video tutorials, a dataset triage tool, and other investigative resources to more than 100 subscribers in a secure environment;
- organised four specialist training sessions in tracing crypto transactions for Data Lab members with a leading blockchain analytics company.

This work surfaced key lessons on how researchers and journalists can and should work together, including:

- Journalists and academic researchers operate within different ethical and legal frameworks, but they can each learn from each other's experience of working with sensitive data to ensure their collaborations serve the greater good and minimise harm.
- The very different timelines for most academic and journalistic projects can make synchronised publication seem unworkable, but playing to their strengths within a flexible and iterative project can produce mutually reinforcing, coordinated articles that maximise impact.
- The different traditional priorities of journalists and academics – telling individual stories vs conducting systematic analysis – can appear to pull in different directions, but there is increasing overlap in the tools each uses to analyse large datasets. Provided collaborations are built upon clearly defined roles and expectations, the skills and expertise of each group can complement and reinforce each other.

Journalists and academics working on illicit finance are also tackling many of the same themes.

On cryptocurrency, the opportunities for greater collaboration are significant. Data Lab activities suggest that cooperation with academics could help journalists follow the many leads they currently have, but are struggling to address due to technical constraints. These interactions, if structured well, would also provide researchers with fresh data on how crypto is being abused, particularly for money laundering purposes.

Professional enablers is another priority area of work. There is a clear need to better assess the research potential of existing datasets on banks, corporate service providers and other professionals, to build a deeper understanding of how these actors support illicit activities.

Participants also began sharing strategies for accessing and analysing property and beneficial ownership registers, such as how to efficiently create PEP lists for cross-checking, and navigating the new legitimate-interest access regimes emerging in Europe and beyond. These strategies now need to be implemented.

Looking ahead, the Data Lab will support the establishment of several working groups on these priority themes, along with other threats such as the scam industry, which has exploded in recent years.

We want to ensure journalists and academics not only have the knowledge and connections to carry out this complex work, but also the raw materials: the right tools and data.

For that reason, OCCRP plans to expand access for academics to its data platform, Aleph Pro, and the Data Lab will curate and share new and underexploited datasets with the community.

Flows of dirty money pose an urgent and highly complex challenge to people, politics and the planet. Responding to it requires a joined-up ecosystem that shares skills, expertise, data and evidence while respecting professional boundaries. The Data Lab is a new but essential part of this landscape.

Introduction

“What is the problem we are **collectively** trying to solve?... Not the one my organization is trying to solve. The **shared** problem. That’s where every project has to start.”

That was the simple but fundamental question put to a room of around 40 leading journalists and researchers specialising in corruption and organised crime at the Illicit Finance Data Lab workshop in Amsterdam in March 2026. The answer was clear: the dirty money flows fueling transnational crime, kleptocracy and other threats to peaceful, democratic societies.

The scale of the problem is staggering. Organised crime is [estimated](#) to account for up to 7% of global GDP. Undeclared offshore wealth is even higher, equivalent to nearly 10% of global GDP or \$13.6 trillion, according to the Tax Justice Network. These figures should not reduce these challenges to abstractions; corruption and economic crime inflict devastating real-world effects which are helping to undermine democracy, the environment and human security across the world. Globalised flows of dirty money sustain these harmful activities, from embezzlement and bribery to scams, trafficking and sanction evasion.

To counter them effectively, the anti-corruption community cannot rely on fixed, siloed systems designed to respond to the challenges of a different century. A shape-shifting problem requires a dynamic, resilient counter-movement.

In the face of these threats to economic growth and public safety, policymakers, law enforcement and private sector players are looking for the most resource-efficient and effective ways to fight back.

Utilising the vast quantities of relevant data is one. Recent years have seen an explosion in the quantity of data that is relevant to uncovering, tracking and understanding illicit financial flows, thanks to leaks, new technologies, and the hard work of transparency campaigners.

Journalists and academic researchers each have skills and approaches essential to unlocking the vast potential of the data available. Their specialities complement each other, enriching investigations and creating new opportunities for impact. These partnerships are an essential ingredient in the collective, all-society approach that is the only way to fight modern-day global kleptocratic networks at scale. But the challenge is not straightforward, as these two groups have traditionally remained separate: “Like oil and water,” as one Data Lab member put it.

That’s why we established the Illicit Finance Data Lab, to unlock new, high-impact collaborations across sectors.

Much of the expertise needed for these kinds of collaborations already exists, but is often trapped in silos, or in the heads of journalists and researchers, who have built up specialist knowledge over many years, sharing their experience primarily with their own professional community. The Illicit Finance Data Lab provides a space for a community of experts to connect and learn from each other.

This report shares valuable findings and insights from the first phase of the Illicit Finance Data Lab, providing important context for why journalist-researcher collaborations are essential in furthering the fight against global financial crime and corruption.

Combining the Data Lab’s own research with extensive input from the anti-corruption community, the report examines the most pressing challenges facing data-led investigative collaborations, spanning legal, ethical, data processing and funding issues, as well as cultural differences, and offers some of the solutions to these challenges identified during the start-up phase.

Later sections examine the complex data landscape facing journalists and researchers, and dig into priority topics of investigation, including cryptocurrency, beneficial ownership and property registers, as well as the professional services enabling corruption.

And finally the report looks forward, setting out the priorities for the next phase of the Data Lab.

Many of the lessons noted here will apply beyond the illicit finance community, and can inform collaborative projects involving scientists and journalists working together on subjects like climate change, health and the environment.

The recommendations in this report are drawn from the Data Lab's discussions and experiences. Due to the emerging nature of the field and the speed at which technology is driving new research techniques (and new tactics in illicit finance), we fully expect the recommendations to change over time. And, in the spirit of collaboration, we are particularly interested to hear from anyone who disagrees with our findings and wants to share a different perspective.

"I came back with more ideas than I can currently handle, which is exactly the right problem to have after a workshop like this"

- Amsterdam workshop participant



I. The Importance of Collaboration in Tackling Financial Crime and Corruption

If we think of global financial crime and corruption as a pandemic, illicit finance is the vector through which the virus is able to spread. To defeat the virus, or at least curb its most harmful effects, we have to be able to diagnose how it moves through the world and what treatments work best.

The growing availability of massive datasets provides journalists and academics an unprecedented opportunity to do precisely this work. The incentive to collaborate is partly driven by the question of resources: many journalism organisations face increasingly limited staffs and budgets. Meanwhile, academics need access to new sources of data to publish new, meaningful research. For both groups, generating real-world impact from their work is not only a personal imperative, but increasingly a fundraising one, as well.

The benefit of collaboration is not just to bring greater manpower to study the vast number of relevant datasets, both public and leaked, that have become available in recent years. The skills and perspectives that each group brings to the table enable them to unlock different valuable insights from the same datasets.

Several projects of OCCRP and ACDC helped inspire the Data Lab, and illustrate how this works in practice (see Examples of Effective Collaboration, below). Following the biggest ever leak of banking data from Africa, journalists exposed how political elites in the Democratic Republic of the Congo had looted the country on a vast scale. Later, academics analysed transaction patterns over time to identify systemic vulnerabilities in the global banking sector, including massive compliance failures from international correspondent banks and auditors that enabled the corruption to go unchecked.

In another example, journalists working with leaked data identified public officials, criminals and sanctioned individuals buying up luxury property in Dubai. Academics asked macro questions, identifying the scale of foreign real estate ownership and the most common nationalities of those buying it, including Russians after the invasion of Ukraine.

In these collaborations, and others summarised below, journalists used the data to identify individuals at the heart of their stories, putting “a human face on corruption” (to borrow a phrase from one Data Lab member). Academics carried out a more extensive analysis of the data to identify patterns and relationships, and to test for causal explanations for the illicit activity over time.

In other words, journalists were looking for needles. Academics were analysing the entire haystack.

Other collaborations demonstrate additional complementarities. Speaking to the technical expertise available in academia, one team of computer scientists developed an AI tool that extracts structured data from a large store of corporate registration documents held by a journalism organisation. The resulting dataset contains companies, their owners and their professional service providers, and can feed into reporting and research on a leading illicit finance jurisdiction.

In other instances, researchers are the ones collecting the evidence or data, and sharing it with journalists when it contains newsworthy findings or can supplement their investigations.

Another advantage of collaborations is that journalism and academia’s different outputs can drive greater impact when combined. So while an academic paper might struggle to reach many readers outside a specific research community, a high-profile journalism investigation, driven by collaboration with the same academic, could engage an audience of millions. With the topic suddenly elevated in the public discourse, the rigorous academic research provides policymakers and enforcement authorities with the systematic evidence needed for changes in policy or enforcement strategies.



II. The Data Lab Response

Despite the clear benefits to both sides, meaningful collaboration between journalists and researchers in this field has been rare up to this point. The Data Lab set out to change that.

Inspired by our own experiences of collaboration, the journey began with a workshop organised by OCCRP, ACDC and the National Endowment for Democracy in London in 2024. There, 35 experts from journalism, academia and civil society discussed ways to better exploit massive datasets for illicit finance investigations, and identified the core needs and challenges facing the community. Participants agreed that alone they lacked the “analytical firepower” needed to leverage the rich datasets now available.

Launched officially in January 2026, the Illicit Finance Data Lab’s three-month start-up phase sought to do three things: bring people together; kickstart collaborations; and fill knowledge and capacity gaps. In the words of one member, “The Data Lab is like the circuit board... it enables us to connect different investigative pathways”.

Forging a Community

One basic barrier to collaboration is that journalists and researchers work in separate professional universes. The Data Lab creates a space for them to interact.

To build our network we reached out to the research community at in-person events and two webinars, co-organised with the Governance & Integrity Anti-Corruption Evidence programme (GI-ACE), a leading research initiative based at the University of Sussex.

Data Lab colleagues interviewed journalists and researchers from 10 countries to better understand the challenges and experiences of collaborations between these communities.

We created an [online platform](#) to share guidance materials, host discussions and provide updates on training sessions and events. Over 100 people have signed up, creating a virtual meeting place for the Data Lab community that will continue beyond this phase of the project.



Participants at Illicit Finance Data Lab workshop in Amsterdam

But virtual interactions are never enough. Many illicit finance investigations cannot be discussed openly online. Practitioners need to meet in person and share ideas and experiences. In an emerging field where collaborations across disciplines are still relatively rare, practitioners don't always know who else is working on similar themes, or what data and techniques they are using that could inform their own work.

Over two days in Amsterdam, we brought together 37 journalists, researchers and data specialists to develop and share best practices for collaboration; explore the research and investigative potential of promising, under-exploited datasets; and help develop the networks and resources needed to support the next generation of journalist-researcher collaborations.

Several attendees noted the importance of the Data Lab in actively fostering community among otherwise disparate groups of researchers and journalists who often lack time and resources to do it themselves.

"You need to have someone playing that centralised role, pushing people together every now and then. Otherwise nobody is actually going to do it."

- Amsterdam workshop participant

Advancing Collaborative Projects

The Data Lab has advanced 11 specific collaborations, uptake that demonstrates the active demand for this type of engagement. The illicit activity or the data involved in these projects comes from more than 15 countries, including leading financial centres and Global South countries. The ongoing work includes:

- compiling and sharing data gathered by academics on over 20,000 PEPs with investigative journalists in several African and Middle Eastern countries;
- compiling data on over 100,000 companies and individuals linked to heavily sanctioned jurisdictions;
- processing beneficial ownership data at scale to make it more usable by journalists, and sharing with partners in 18 countries;
- securing researcher inputs on how best to interrogate new stores of property and beneficial ownership data recently acquired by media partners so that the eventual reporting contains actionable analysis;
- testing new big-data methods for identifying sanction evasion.

Filling Knowledge & Capacity Gaps

Our engagement with journalists and researchers made clear that certain knowledge gaps were holding back their work. So we produced a set of initial resources to help overcome obstacles.

Familiarity with a dataset is essential to defining roles and managing expectations ahead of a possible project. Our template for conducting [Dataset "Triage"](#) will help researchers and journalists establish the risks and limitations of datasets at the initial planning phase of a project, along with an accompanying [guidance note](#).

To increase awareness of success stories and inform future project design, the Data Lab compiled summaries of 24 diverse collaborations between journalists and researchers.

A collection of original guidance articles draw together best practice, literature reviews and case studies on collaborative approaches. These were informed by 15 interviews with journalists and academics.

- [Overcoming Cultural Barriers to Collaboration](#)
- [Navigating Ethical and Legal Challenges in Investigative Collaborations](#)
- [Bridging Data Processing Divides between Journalists and Academics](#)
- [Cryptocurrency and Illicit Finance](#)
- [Resourcing Investigative Collaborations](#)

Two public webinars in collaboration with the Governance & Integrity Anti-Corruption Evidence programme introduced the Data Lab, its themes, key members and offerings to the wider research community.

GI ACE SEMINAR SERIES IN ASSOCIATION WITH THE ILLICIT FINANCE DATA LAB

2-3PM GMT | 9-10AM ET | 3-4PM CET
THURSDAY 5 MARCH

SOURCING DATA FOR THE NEXT GENERATION OF ILLICIT FINANCE RESEARCH

 Liz David-Barrett (Chair) Programs Director GI ACE Director Centre for the Study of Corruption	 Michelle Kandler-Kretsch Research Fellow Anti-Corruption Data Collective	 Alexandra Gillies Chief Impact Officer Organized Crime and Corruption Reporting Project	 Amra Dzonlic Zlatarevic Research & Data Operations Manager Organized Crime and Corruption Reporting Project	 Milo Z. Trujillo Postdoctoral Researcher Network Science Institute at Northwestern University
--	--	--	---	---

ACE GOVERNANCE & INTEGRITY Anti-Corruption Evidence Research Programme

UK International Development Partnership | Progress | Prosperity

US UNIVERSITY OF SUSSEX

CSC COLLABORATION FOR SUSTAINABLE COMMUNITIES

ANTI-CORRUPTION AND INTEGRITY

Organized Crime and Corruption Reporting Project

Watch the webinars [here](#) and [here](#).

Cryptocurrency quickly emerged as an area where Data Lab participants sought to level up their capacity. So, the Data Lab worked with [Crystal Intelligence](#), a blockchain analytics firm, to provide several training sessions. The OCCRP network, which includes journalists in dozens of countries, participated in two sessions, while another mixed group of almost 30 investigators and researchers took part in a similar online session. To go deeper, 11 journalists spent a full day with two trainers, digging through actual cases of suspected crypto money laundering and further developing their skills.

We also produced a training video to guide journalists and researchers without coding experience on how to use AI to scrape data from public websites. With such a resource, a researcher or journalist can collect the data they need for an illicit finance investigation themselves, allowing data scientists and programmers to contribute their skills where they are most needed.



Watch the training video [here](#).

Finally, the Data Lab has initiated a pro-bono legal request with an international law firm to assess the legal risks that researchers working with hacked or leaked data may face, and mitigation solutions.



III.

Navigating Challenges to Collaboration

Collaborations between journalists and academics can make for richer investigations that have greater potential for impact, but it's important to be clear-eyed about the significant challenges they entail.

Drawing on all the work outlined above, these are some of the most common challenges encountered when trying to work together.

Ethical Frameworks

Academic researchers have traditionally been required to obtain informed consent from anyone involved in their research before they begin, a principle rooted in the landmark 1979 [Belmont Report](#) in the US, which established core ethical standards for research involving human subjects. This is different for journalists, who typically work on an investigation first, then offer subjects a right of reply before publication.

This disparity can be especially difficult when working with large leaked datasets. Requiring an academic to seek consent from thousands of people included in something like the Panama Papers would obviously be impractical, as would emailing an oligarch suspected of corruption to ask if they minded having their leaked tax documents analysed.

Academics must also demonstrate **public benefit** before their research even starts. Investigative journalists rely on the **public interest** defence to justify revealing information about individuals. Despite the similar-sounding terms, these are very different frameworks. Public interest protects journalists from legal claims arising from publication, while public benefit requires academics to prove, in advance, that the value of the research outweighs any potential harm to those whose data is being used.

In many academic disciplines, it is common to submit research proposals to an Ethics Review Board (ERB, also known as Institutional Review Board, or IRB) before starting work. This can be slow and demanding, potentially frustrating collaboration with journalists working to tight deadlines. Such boards usually draw from different disciplines, so that philosophers or chemists with no expertise in illicit finance can be called on to judge the public benefit of research into offshore wealth or corrupt contracting.

Another important ethical issue is the protection of sources. Public universities are legally obliged to respond to Freedom of Information Requests. This could include the details of an ethics review board application, which could potentially expose a whistleblower who leaked data about dangerous criminal groups.

Potential solutions

Journalists may need to allow for longer timelines and be prepared to make a strong case for why a project has genuine public value. Academics, in turn, should recognise that working alongside journalists can actually strengthen their ethics applications, since collaboration with a newsroom adds a clear, public-facing dimension to the research.

Newer frameworks like those contained in the [Menlo Report](#) (2012) and the DPIA (Data Protection Impact Assessment, now legally required in the [EU](#)) offer practical tools for planning responsible data use, including in cases where consent is impossible. Ethics boards are also becoming more receptive to research using leaked data, particularly in the wake of high-profile investigations like the [Panama](#) and [Pandora Papers](#) and a growing body of research drawing on the related [Offshore Leaks](#) dataset available online.

“Ethics are not rules, they’re principles”, argued one academic speaker at the Amsterdam workshop, who made a passionate case for universities to be flexible and pragmatic in how they apply ethics policies to the illicit finance field, especially given that researchers were often far more vulnerable than the corrupt individuals they were investigating.

Their advice was for academics to “tackle ethics head-on; don’t put your head in the sand.” They argued that because illicit finance investigations often involve serious harms to society, it should be possible to make a convincing case to an ethics or institutional review board that the public benefit of the research outweighs the traditional research need for informed consent. To do this, researchers should draw on ethical frameworks from across multiple disciplines, including journalism, to build a comprehensive case. One example would be for academics to adopt the journalistic principle of a “right of reply” for those affected by the investigation in cases where requesting informed consent is impractical or self-defeating.

In addition to permission, academics that invest in the process are more likely to receive proper institutional backing in the event that legal or security threats materialise in the course of their research.

Legal Risks

Journalists benefit from public interest exemptions that often, but not always, shield them from data protection and privacy laws when working on stories that serve the public good.

For academics, those exemptions don’t automatically apply and university legal teams often lack the specialist knowledge on the specific challenges raised by sensitive data investigations. This was the experience shared with the Data Lab by one professor, who said that research teams at his university were left to deal with complex legal queries on their own. Another said they were worried about “the personal liability of using leaks” given the reluctance of their institution to provide adequate support in the event of legal action.

Further complications arise when data is shared between collaborators. Even publicly available datasets can carry restrictions on how they are used or who can access them. Sharing data across those boundaries, between a journalist and an academic, for example, can create unexpected legal exposure for both parties.

Handling data that was originally obtained through hacking or leaking can create legal problems even if neither party was responsible for obtaining it. Data protection laws (GDPR) around personal information are stricter in the UK and Europe than in the US, meaning that there are further jurisdictional challenges depending on where the research is being carried out, especially if a collaboration takes place with teams in different countries.

Potential solutions

Specialist legal advice is essential – if necessary, in addition to review by a university's in-house lawyers. Organisations like the International Lawyers Project offer pro-bono support to researchers and journalists working in the public interest.

Data Lab members suggested that academics could reduce legal risks by allowing journalists, who enjoy greater legal protection under the public interest defence, to handle the most high-risk data in a collaboration, though the practical impacts of this approach on research require careful consideration.

Another solution shared for sensitive datasets was to only access them on someone else's server (e.g. OCCRP's Aleph Pro database) so that the researcher doesn't need to download the data onto their own system. However, this workaround is not always practical for quantitative researchers who use their own specialist statistical software to analyse data.

The view of one lawyer who attended the Amsterdam workshop was that although the risks should not be disregarded, criminal prosecutions relating to the use of leaked data have overwhelmingly focused on the hackers, not the academics and journalists who may have used the data in investigations.

What's above should not be confused with the legal risks associated with so-called SLAPPs (Strategic Lawsuits Against Public Participation) and other types of "lawfare". Spurious legal claims brought by wealthy individuals to stop the exposure of wrongdoing pose a significant risk to investigations targeting wealthy figures, which should be mitigated through taking specialist legal advice.

The Data Lab has initiated a pro-bono legal request with an international law firm to officially assess the liabilities of academics using leaked data.

There are a number of valuable academic [resources](#) for navigating both the legal and ethical challenges around using leaked data, which are included in this Data Lab [article](#) on the issue.

Differing Timelines

Timelines are one of the most clear-cut differences between journalism and academia and can be a point of tension in collaborations.

The distinction can be stark. As one journalist with extensive experience collaborating with researchers told the Data Lab: “Journalists cannot spend two years researching the same thing.”

These differing timelines can result in missed opportunities for academic research to reach a wider audience. One researcher expressed disappointment that this exact issue meant they missed out on media coverage of research related to a high-profile leak of US law enforcement data:

“Because the academic study took so long, about a year to do all of this analysis, the media focus had died off and moved on by that point”.

Potential solutions

Several Data Lab members affirmed that setting clear expectations about timelines at the start of a project had enabled them to build successful collaborations without any issues. Academic outputs can be released after journalism articles, even if it was the academic analysis of a dataset that surfaced the investigative leads.

When initial findings are robust enough, academic working papers, case studies, research notes or policy briefs can be released in sync with a journalistic investigation, with fuller papers that go through the lengthy peer review process being published later.

Research Approaches

Journalists approaching a large dataset will often have specific targets or red flags in mind and will look for evidence of individual wrongdoing, for example searching a dataset to see if they find matches for politically exposed persons (PEPs).

This approach is less relevant for academics conducting quantitative research on a dataset, who will be more interested in carrying out a systemic analysis, studying the entire dataset to establish baselines, patterns and macro trends backed by a more rigorous methodology.

One Data Lab member described the challenge like this:

“Academics want to have the question clearly defined before they begin any of the analysis... you make the hypothesis and then you test it. And that really does not work well in leaked datasets where you have to do a lot of exploratory work up front to understand what’s in that data.”

Potential solutions

In any collaboration, it is important to share a common understanding of how different skills, expertise, methodologies and technical capacities contribute to a shared objective. In the types of collaborations we are discussing, it is likely that each side will produce mutually reinforcing, complementary but distinct findings. At the risk of oversimplifying, journalists produce the cases that illustrate the trend analysed by academics. Collaborations can run into difficulties when practitioners attempt to delegate research for “micro” findings to someone applying methods better suited to “macro” analysis. In a true collaboration, neither side is there only to serve the other but can contribute meaningfully on their own terms.

Funding challenges for collaborations

Resourcing interdisciplinary collaborations between journalists and researchers is complex because the two fields operate under fundamentally different financial structures. Journalism organisations work with tight editorial budgets, while researchers depend on short- to medium-term grants tied to specific projects and agendas. Informal arrangements leave researchers, particularly those at an early-career stage, vulnerable to exploitation and scope creep, with no guaranteed access to institutional resources.

Hiring academic collaborators as consultants places the full financial burden on newsrooms, which is often unrealistic. Meanwhile, professional norms in academia mean that unconventional funding sources can carry a degree of stigma, and analysis conducted outside formal institutional processes may be ineligible for academic publication.

Potential solutions

Several models can address these challenges. Researchers may draw on existing outreach or knowledge-transfer budgets within their grants to work with journalists, keeping the two parties institutionally independent while legitimising the collaboration. University-managed consulting frameworks offer a more structured paid arrangement while preserving institutional cover.

For larger, longer-term projects, joint grant applications represent the most sustainable solution, enabling both parties to enter the collaboration as equal, funded partners with clearly defined roles, governance structures and genuine institutional backing.

[Read the full article on the Data Lab website.](#)

Examples of Effective Collaboration

Journalists and researchers have found innovative ways to collaborate on data projects, and these projects provide inspiration and guidance for expanding this type of work.

Projects serving as direct precursors to the Data Lab include:

Identifying Enabler Networks and their Vulnerabilities

Journalist role: OCCRP supports researchers to access and explore its [Aleph Pro](#) database.

Researcher role: Political scientists from Universities of Exeter and Oxford draw on this data to [map the networks](#) of professional enablers servicing politically exposed persons from high-risk countries. Computer scientists from Exeter also developed an AI model that extracted structured corporate service provider data from a large store of corporate filings held by OCCRP. This data will inform future journalism.

Congo Hold Up: How a Captured Bank Enabled Grand Corruption

Journalist role: The Platform to Protect Whistleblowers in Africa (PPLAAF) received a 3.5 million document leak from the BGFIBank in the Democratic Republic of Congo (DRC). After journalists published [dozens of stories](#), PPLAAF shared and jointly explored the data with [researchers](#).

Researcher role: Academics working with [ACDC](#) – including researchers from George Washington University, Copenhagen Business School and the International Tax Observatory – analysed the data to evaluate the impact of key regulatory policies on illicit transactions, and have identified new money laundering schemes through the bank in the process.

The above two projects received funding from the UK's Anti-Corruption Evidence program, a crucial source of support for innovative collaborative research.

Dubai Property Ownership

This work relied on data acquired by Center for Advanced Defense Studies (C4ADS), a NGO, who also [examined](#) the data for evidence of sanctions evasion.

Journalist role: Dozens of media organisations, coordinated by OCCRP and E24, worked together to identify and [report on](#) hundreds of public officials, criminals, and sanctioned individuals who held Dubai property.

Researcher role: Academics from the International Tax Observatory [estimated](#) the value of foreign-owned real estate, analysed country-level ownership patterns, and examined whether property owners declared those assets to their home tax authorities.

Secretive Firms From UK to Singapore Win \$200M in Tenders From Uzbek State Giant

Journalist role: OCCRP and Finance Uncovered journalists [reported](#) extensively on the contracts awarded by a large state-owned Uzbek mining company, including to UK companies exhibiting multiple red flags of suspicious behaviour.

Researcher role: A professor from Ulster University first compiled and analysed the company's tender data and shared the findings with journalists, kickstarting the project. His research on the case is ongoing.

Other projects – on illicit finance and very different topics as well – further illustrate how these collaborations can work.

Offshore Leaks and FinCEN Files

Journalist role: The International Consortium of Investigative Journalists (ICIJ) led the collaboration of journalists that published history-making reporting on the Pandora Papers, Paradise Papers, Bahamas Leaks, Panama Papers and FinCen leaks. They then made much of this data publicly available, in the [Offshore Leaks Database](#), which includes 810,000 offshore entities, and the [FinCEN Files](#).

Researcher role: The data published by ICIJ has informed a rich set of research [publications](#) on tax evasion and related topics.

DATA-CROS: Developing Tools to Assess Corruption, Collusion and Money Laundering

Journalist role: Journalists feed into and use the tools. The Investigative Reporting Project Italy (IRPI) drew on the tool to report on [sanction evasion](#) and [organised crime](#) activities in Rome.

Researcher role: Researchers from the Università Cattolica del Sacro Cuore coordinate the data tool's development and conduct analysis, including of corporate [ownership anomalies](#) across 29 European countries.

Relationscope: Social Network Analysis of Slovenian Politicians

Journalist role: Oštro developed an interactive [data tool](#) that allows the public to find links between Slovenian ministers, secretaries of state and parliamentarians along their education and career paths. Along with gathering the data, Oštro [published](#) stories on the importance of university ties and mapping the background of new politicians.

Researcher role: Academic experts in statistics and mathematics provided support, including the application of specialised large-data social network approaches.

Deaths in the Family: An Investigation into Transgender Homicides

Journalist role: Journalists from Business Insider collected data on 175 cases, interviewed experts and [wrote](#) stories on the deaths and their aftermath – including finding that more than one-third of the crimes remain unsolved.

Researcher role: A computer scientist supported the data analysis, and a gender and sexuality expert provided sensitivity reads of story drafts. The data also formed part of a PhD thesis and an academic [article](#).

How Tiktok Hijacks Politics

Journalist role: Journalists from VSquare, Investigace and other Central European outlets examined how disinformation and propaganda spread across video social media platforms. Their stories tackled TikTok's [algorithm](#) and its effects, and political events in [Romania](#), [Slovakia](#), and [Czechia](#).

Researcher role: A data scientist and disinformation expert from Charles University (Prague) developed the tooling needed to analyse vast quantities of video content.



IV.

Data: Navigating a Complex Landscape

The telltale signs of illicit financial flows are often hidden within a maze of complex data, accompanied by thorny issues concerning access, processing and collaboration.

A single data leak can be overwhelming in scale, with some of the largest cross-border journalistic collaborations examining millions of files in multiple formats, often including spreadsheets and other files that themselves contain hundreds or thousands of records. Information about who holds or has access to leaks is usually carefully controlled, not only to minimise scoop risks for journalists, but to protect whistleblowers and other sources.

Public datasets such as beneficial ownership registers and land registers easily run into millions of entries and can contain a bewildering array of fields. Manual searches can miss critical details, and barely scrape the surface of the information available. These datasets sometimes contain layers of information in both web pages and the underlying documentation available for download. On top of this, shifting access regimes, uneven data fields, formats and structures can be a barrier to cross-border investigative efforts.

Commercially available data can be some of the most valuable sources for understanding the economic activity connected to illicit financial flows. These data tend to be well-structured, pre-processed and validated, although online dashboard interfaces and licensing agreements may preclude the kinds of additional processing and matching favoured by researchers and investigative journalists alike. Financial barriers and uneven distribution across the anti-illicit finance community can pose challenges to effective cooperation. On the other hand, no- or low-cost licenses available to researchers can create an opportunity for collaboration with journalists.

Advanced data science and academic research methods can help make sense of all these categories of data, accelerating accountability and informing policymaking.

The Data Lab has begun building an inventory of public and leaked datasets that could form the basis of collaborative projects. Some of these have previously been used by journalistic organisations, and researchers may be able to extract more value from the information they contain. Others have not yet been comprehensively examined by anyone, and the insights generated by academic analysis can unlock findings and indicate directions for further investigation.

Working with the Data Lab, **Distributed Denial of Secrets (DDOS)**, a US non-profit that archives and publishes hacked and leaked data in the public interest, revisited its extensive inventory to propose a series of datasets underutilised by academic researchers of illicit finance.

OCCRP has published a selection of promising datasets from the Aleph Pro library, including detailed summaries covering the entities, structure, geography, time period, languages and other details that can help inform potential uses. Data Lab activities have informed new thinking at OCCRP about how to make Aleph Pro more useful to researchers, and those ideas will be implemented in the coming months.

ACDC has applied methods informed by academia and data science to process several public datasets on asset and company ownership and shared the results with journalists in relevant countries, forming the basis for collaborative investigations with potential to both expose cases, build evidence on how illicit financial flows occur and assess the effectiveness of policy and enforcement solutions.

A blurred image of a busy city street with a digital circuit overlay. The image is dark and has a greenish tint, with a complex pattern of white and yellow lines resembling a circuit board or data flow overlaid on the scene. The background shows people walking and buildings, but they are out of focus.

V.

Leading Themes

Journalists and researchers consulted by the Data Lab identified several priority topics for high-impact analysis. This has guided our approach to dataset selection, training, guidance notes, and helped set the thematic focus for the working groups that will form part of the Data Lab's next phase. Along with the themes detailed below, there is also clear demand for cooperation on sanctions evasion, disinformation and scam networks.

Cryptocurrency

Researchers and journalists are among the many anti-corruption players scrambling to up their game on cryptocurrency, as it balloons as a vehicle for illicit transactions.

Investigations into crypto-related crime have typically focused on four high-risk areas: regulated exchanges that facilitate money laundering; informal cash-to-crypto swaps involving “money mules;” the fraudulent issuance of new coins (including high-profile “meme coins”); and large-scale scams like “pig-butcher” operations in Southeast Asia. These activities fund everything from paramilitary operations to luxury real-estate fraud.

To help researchers and journalists bridge the gap between anonymous digital wallets and the individuals controlling them, the Data Lab [has outlined three primary investigative strategies.](#)

- Contracting with specialised commercial firms. While these companies possess powerful data and proprietary tools, the guide warns of high costs and potential conflicts of interest.
- Partnering with academic experts. This path promotes open-source methodologies and sustainable knowledge transfer, though may require aligning the slower timelines of academia with the fast-paced needs of journalism. Success in this area often depends on providing researchers with unique, unexploited datasets.
- Building in-house capacity using “off-the-shelf” tools and APIs. While this requires resources for hiring, coding expertise and initial setup costs, it offers the greatest long-term flexibility and independence.

The Data Lab provided four crypto training sessions in March 2026 in partnership with a leading crypto analytic firm. At one, an all-day session with a small group of journalists, the participants worked with leaked data from an active investigation to tie wallet addresses to suspicious activity in ways that can then be verified and fact-checked.

Initial networking between journalists and leading academic experts also proved promising, and can hopefully help journalists to depend less on commercial firms alone. The working group will advance these relationships, and also survey leaked data on crypto from DDOS, OCCRP and other sources.

Beneficial Ownership and Property Data

Property and corporate ownership registries are among the most valuable sources of information for journalism and research on illicit financial flows. The first generation of investigations and research on this data has produced valuable insights. Examples include the work of journalists and academics on the property and corporate registries of the UK and France, and leaked [Dubai property data](#)¹.

Both journalists and academics want to build on this foundation and extract more public interest value from these datasets. Learning from each other's methods will help. For instance, in Data Lab discussions, journalists discussed how they tend to apply the same methods to these datasets, scanning them for "bad guys," but that they want to ask more macro questions about trends and tactics. Researchers can add great value here, in helping to formulate these questions and advising on approaches. Options include looking for clusters and patterns, such as companies formed on the same day, to generate leads, and examining how global events and policy changes (e.g., new tax rules) drive people from different countries to set up offshore companies.

Towards this end, several ongoing projects related to specific property or beneficial ownership registries are underway. Some feature researchers advising journalists on new ways to interrogate a newly available dataset, which they may also use for new research. Others have researchers helping media outlets in Africa and the Middle East to develop the lists of politically exposed persons needed to spot possible illicitly held assets in foreign property data.

Collaboration could also help both groups to navigate the very complex and shifting landscape of data availability, including the new legitimate interest access regimes emerging in Europe and beyond. Just keeping track of which registers are available and how to access them can be overwhelming. The Data Lab will coordinate with leading transparency groups to make this information available to researchers and journalists, beginning with the status of property disclosures across the G20, and explore how to share register access across organisations within the various regulatory constraints.

1. For a collection of research on offshore real estate ownership by location, see: <https://www.taxobservatory.eu/repository/offshore-ownership-of-real-estate/>

Professional Enablers

Another hot topic is the role played by banks, corporate service providers, law firms and other professionals in executing corruption and money laundering schemes. To understand their activities, researchers are conducting everything from ethnographic deep dives to big-data analysis of leaked banking transactions.

Journalists have received many leaks of internal files belonging to these players, and have reported widely on this unique behind-the-scenes material. Examples include the Cyprus Confidential project and OCCRP reporting on Credit Suisse and several Swiss banks.

According to Data Lab participants, these leaks may have research potential as well. Structured financial data can empower data analysis. Qualitative researchers could use emails and other text files to understand the relations between these firms and their clients, and how they manage (or ignore) the risks associated with handling suspicious activity.

One major leak covered a dozen years, so could reveal how the services changed over time. Across these possible uses, determining data completeness is a challenge which researchers would need to manage.

Banks are a priority. The global financial system puts great trust in these institutions to police their transactions for suspicious activity. So when they are captured or actually purchased by a particular interest group, the effects can be quite serious. ACDC and PPLAAF collaborated to analyse a leak of banking data from a captured Congolese bank. Examining ownership changes of banks or other banking leaks could build on this work.

Conclusion

This initial phase of the Illicit Finance Data Lab has allowed us to establish a collaborative community of practice on illicit finance, addressing pressing gaps in knowledge and capacity, sharing data, and kickstarting new collaborations which will yield new learnings and deliver real-world impact.

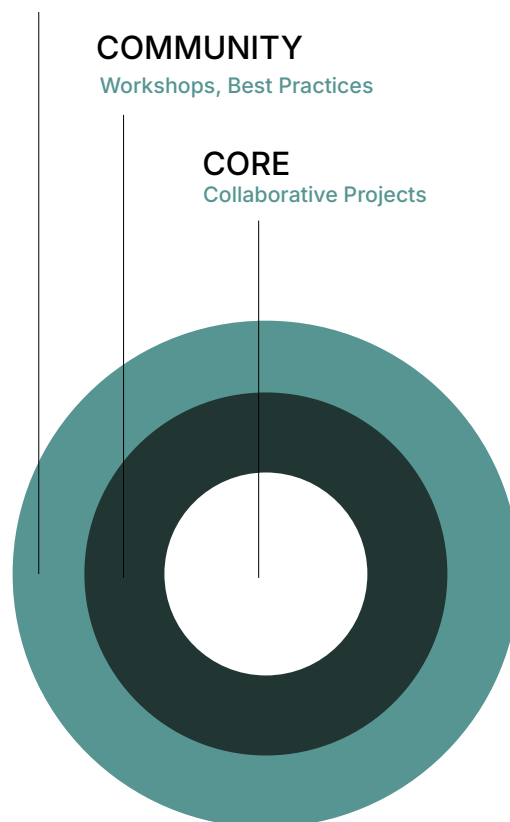
Future success in this field will depend on both continued engagement from the Data Lab, and momentum from the community. That's why we are creating thematic working groups to allow independent peer-to-peer exchange and cooperation that can sustain the results of this phase. The Data Lab team will continue to facilitate and match-make within and between these groups, develop the collaborative projects initiated so far, and share insights with the community as we go.

But there is a long way still to go. We recognise that the level of experience and comfort with the kinds of collaborations we are promoting continues to vary widely among both journalists and academics. Moving forward, we propose a three-pronged approach to:

- **support a core group** to execute projects through data provision, matchmaking and thematic and project workshops.
- **build the community** with access to dedicated resources, best practices and peer-to-peer exchange, with opportunities to collaborate and learn through working groups, the online platform, and broader convenings.
- **lower barriers to entry across the ecosystem** with public communications and partnerships with academic and media networks to help build familiarity and overcome the initial hurdles faced by others.

ECOSYSTEM

Lowering Barriers to Entry



In this system, the lessons learned through practice project outwards to inform the community and the broader ecosystem, creating positive feedback loops that bring more researchers and journalists into the community and lead to more projects from which we can all learn.

Ultimately, a thriving ecosystem will be self-sustaining. At this stage, real progress towards this depends largely on our ability to foster new, valuable, collaborative projects with powerful demonstration effects. This in turn, depends on a critical mass of academics and journalists, including and beyond ACDC and OCCRP's immediate network, having the knowledge, familiarity, support and confidence to undertake projects together – precisely the community we have set out to create during this phase.

Having sown the seeds in the past months, in the next chapter of the Illicit Finance Data Lab, we want to help a thousand flowers bloom.



To find out more, or to request access to the Illicit Finance Data Lab web platform, please write to team@acdatacollective.org or info@occrp.org