

# The Transparathon

For Transparency, Accountability and a Better Future

Bringing Together Scholars, Researchers, Journalists, Civil  
Society and Programmers for a Greater Cause

## Table of Content

<b>Introduction.....</b>	<b>2</b>
<b>Possible Topic 1: Transparency in the Real Estate Sector.....</b>	<b>5</b>
<b>Possible Topic 2: Transparency in Luxury Good Markets.....</b>	<b>7</b>
<b>Possible Topic 3: Lifestyle Monitoring.....</b>	<b>8</b>
<b>Possible Topic 4: Monitoring Public Official’s Conflicts of Interest.....</b>	<b>9</b>
<b>Possible Topic 5: Analyzing Company Ownership and Litigation Data.....</b>	<b>10</b>
<b>Possible Topic 6: Spotting Red Flags in Public Procurement.....</b>	<b>12</b>
<b>Possible Topic 7: Spotting Red Flags in the Extractive Sector.....</b>	<b>13</b>
<b>Possible Topic 8: Identifying Individuals and Firms Closely Tied to Senior Government Officials</b>	
<b>Possible Topic 9: Testing Corporate Registration Agents’ Compliance with Verification and Transparency Rules.....</b>	<b>15</b>
<b>The Transparathon-Team.....</b>	<b>16</b>

Corruption thrives in secrecy. Kleptocrats, plutocrats, and the network of unscrupulous intermediaries who facilitate their plunder fear public exposure, which is why “sunlight” and “transparency” have become the rallying cries for those dedicated to fighting corruption around the world. But notwithstanding diligent investigative efforts by law enforcement, journalists, civil society activists, and researchers, much of the information necessary to expose the corrupt remains hidden. Sometimes the necessary information is so well concealed that it will never be discovered without a leak or a subpoena. But sometimes vital information may be hiding in plain sight: the data is out there, but there is so much of it, and it is so disbursed or hard to access, that even the most diligent journalists, researchers, and public officers are unable to make use of it, or even realize that it exists. It is here that cutting-edge technology, and those who know how to use it, can make a huge impact, through relatively simply technical “hacks.”

A case in point: Some years ago, the programmer Daniel O’Huiginn [scraped](#) the official Panama Registry of Companies and made it searchable by an individual’s name. Thanks to this hack, activists and journalists gained insight into how specific high-level corrupt actors laundered their ill-gotten gains using Panamanian shell companies; some of those companies were also shown to serve as slush funds to pay “facilitators” to help launder reputations of murderous kleptocrats and lobby governments to take policy actions that protect their interests. Another example: In 2016, the Swiss journalist Francois Pilet created a twitter-bot, [GVA Dictator Alert](#), to inform the public as soon as the plane of an autocratic and kleptocratic regime leader lands in Geneva. This caused a huge debate in Switzerland about the use of the country to launder money or evade taxes.

Examples such as these are encouraging, but few and far between. Part of the problem is the lack of sustained personal and institutional ties between, on the one hand, the individuals and groups with expertise on the substantive problem (large-scale corruption and associated unlawful or unethical behavior) and, on the other hand, the individuals and groups with the technical expertise necessary to develop the kinds of “hacks” that would have the greatest impact. While some collaboration between the “policy wonks” and the “techies” already exists, more work is needed to develop an effective long-term collaboration between activists, journalists, researchers, and programmers to generate sustainable and effective technical tools and techniques for promoting the transparency that kleptocrats and their cronies fear.

To lay the foundation for this collaboration, we propose to organize a “Transparathon” event in May 2019 that will bring together anticorruption policy experts and skilled programmers interested in collaborating

on developing anti-kleptocracy-tech tools. The solutions developed through the Transparathon and follow-up efforts will be open-sourced and therefore reproducible, meaning that all participants (and others) will be able to use the tools developed in their work.

The proposed Transparathon is inspired by the “hackathons” that have taken place over the last two decades, but it differs in three important ways. First, while most hackathons focus on a single problem identified at the outset, part of the goal of the Transparathon will be problem identification. Second, while at most hackathons the goal of the programming teams is to have a finished product (or at least a prototype) by the end of the event, after which the teams disburse, the Transparathon aims to develop a longer-term collaboration between the technical teams and the civil society organizations, in order to foster sustainability and enable the pursuit of more ambitious projects. Third, while many hackathons are competitive, the Transparathon will be a collaborative enterprise; rather than picking winners and giving prizes, the Transparathon aims to connect programmers with a passion for promoting transparency and global justice with organizations and institutions that can help them to maximize the impact that their technical skills can have on solving real-world problems.

The Transparathon, in short, will not be a one-off event but rather a springboard for ongoing interactions among stakeholders—the “techies” and the “wonks”—in order to create a more sustained alliance of skills and actors to better leverage data and technology for a better world with less corruption, less kleptocrats, but more transparency.

The May 2019 Transparathon will be a two or three day event, organized into working groups committed to devising solutions to a specific corruption “problem statement.” The problem statement will be used to center and focus each team’s efforts to develop appropriate solutions. The problem statement must therefore be both specific enough to provide sufficient focus, but also flexible enough to allow teams to be creative and to take their project in different directions, including directions not foreseen by the project organizers. Each problem statement should consist of (1) a description of the underlying problem (usually some manifestation of insufficient transparency), and why that problem is important to the fight against corruption; (2) a discussion of the kind of information that might be useful to address the problem, including existing datasets or sources that might be useful if the right information could be extracted; (3) an initial sketch of the kinds of technological tools that activists, journalists, policymakers, and others might find most useful.

We are now in the process of developing the problem statements that we will present to the participants in the Transparathon, and we are reaching out to a broad range of stakeholders and potential participants and collaborators for input. The remainder of this memorandum sketches several possible problem statements, on which we invite feedback, suggestions, and constructive criticism. We also welcome additional proposals.

## Possible Topic 1: Transparency in the Real Estate Sector

**The Problem:** Kleptocrats and other criminals often launder ill-gotten assets through real estate. Luxury real estate purchases are also often a favorite way for the corrupt to spend their money—especially in desirable locations like New York, Paris, Miami, London, and Hong Kong. The use of anonymous LLCs is a particularly popular vehicle for corrupt actors to funnel their cash into real estate. Identifying real estate owned by high-level public officials and their cronies can be an effective way of exposing potential wrongdoing in individual cases and pushing for reforms on a broader scale. Unfortunately, in many places real estate and property registers are non-existent or inaccessible.

**Possibly Relevant Information:** Jurisdictions vary widely with respect to the information they make available about real estate ownership and transactions. Some countries (or states/provinces or municipalities) already have public registers of real estate transactions that are accessible in machine-readable form. Some data is available in, for example, the [UK](#), [Spain](#), [France](#), and the [US \(in some states\)](#), but this information is not always used most effectively. In other jurisdictions, the necessary information is even harder to access. Some private firms, such as Zillow and First America Data Tree, collect information on real estate transactions, but restrict and/or charge for access to this data.

**What Tech Tools Might Help?:** One possible solution could be a webpage which maps the level of beneficial ownership transparency in the real estate sector for cities, regions and countries all around the world. For those cities/regions/countries, which already have a public database, the webpage could provide the respective links. For those cities/regions/countries, where the information is available but not in a way that is easily searchable, a scraping-effort might be worthwhile or an effort to harmonize and organize the material in a user friendly way for the general public. Furthermore the webpage could provide something similar to the [FOIA machine](#) – helping users to ask authorities for the relevant data and at the same time making sure, that the data is made public if the petitioner is successful (to prevent users being forced to pay for data others already had paid for). Mapping the places with and without cadasters or registries would also visualize the gaps and challenges we have, provide fuel for calls for better data collection, publication and organization of gradually richer bunches of data, and show the progress made overtime as more information is gather, systematized, analyzed, applied for policy and social action or justice, asset recovery etc..

Another potentially powerful tool to be developed involves the creation of real-time predictive

algorithm so that journalists and citizens can quickly identify real estate transactions that are flagged as suspicious. The predictive tool would work in the following manner:

- 1) Collect data on real estate transactions on a regularly (daily, weekly, or monthly) basis for a given jurisdiction, such as the state of Florida.
- 2) Parse the basic attributes of each sale through a machine-learning algorithm that identifies certain criteria that are potentially strongly correlated with money laundering:
  - a. Run the names of LLCs (entered as Buyers or Sellers) through databases such as OpenCorporates. Look for flags, such as: multiple LLCs registered at one address; LLCs registered under a single agent or nonexistent agent; LLCs registered at false addresses; LLCs linked to money laundering prosecutions
  - b. Identify all-cash purchases
  - c. Use selling-price data to identify quickly flipped units at high mark-up rates

The outputs from such an algorithm could include:

- 1) Regularly updated heat maps (or dashboards) that allow the general public to identify where LLCs are being used to buy up property. This could help policymakers connect the dots between money laundering and decreasing affordability in real estate markets.
- 2) Search functions for journalists to run the names of LLCs or individuals to identify specific real estate transactions and/or holdings, with rankings of suspiciousness

## Possible Topic 2: Transparency in Luxury Goods Markets

**The Problem:** In addition to real estate, luxury goods—art, jewelry, yachts, high-end automobiles, private planes, race horses, etc.—are an attractive way for corrupt officials to launder their illicit wealth, and one of the most popular ways for such officials to spend their ill-gotten cash.

**Possibly Relevant Information:** The amount of data on luxury good transactions varies widely by sector. For example, the art market is largely unregulated, while yacht purchases are somewhat easier to track given that many countries make public data on boat ownership and sales (for example, the [United States](#) and [Canada](#)), as well as [private firms](#).

**What Tech Tools Might Help?:** There are great opportunities to build registries for both goods and fill them out using a variety of publicly available data. We could take advantage of the partial centralization through auction houses and even brokers to collect and process records. Tools could be developed to identify purchases of yachts by public officials, and then trace their travels around the globe. Many companies provide real-time tracking of vessels for a nominal price, including FleetMon (<https://www.fleetmon.com/vessels/>) and VesselFinder (<https://www.vesselfinder.com/vessels>)



## Possible Topic 3: Lifestyle Monitoring

**The Problem:** One of the ways to catch corrupt public officials is to identify a mismatch between the official's legitimate income sources, on the one hand, and his or her assets/lifestyle, on the other. In some countries, such a mismatch can be the basis for a criminal prosecution (for so-called "illicit enrichment") or civil forfeiture of assets (through so-called "unexplained wealth orders"). Yet these mechanisms have not been as successful as their proponents had hoped, in part because it can be quite hard to document such cases (substantial increase in the net worth of politicians and administrators way beyond their declared or official incomes).

**Possibly Relevant Information:** Monitoring and documenting the lifestyle and assets of local or national politicians would require having local eyes and ears, "reading the local newspaper" and observing the routines of suspected corrupt officials.

**What Tech Tools Might Help?:** Tools to identify those living beyond their means and abusing their powers. Furthermore, tools to visualize corruption to strengthen accountability

## Possible Topic 4: Monitoring Public Officials' Conflicts of Interest

**The Problem:** Income and asset disclosures by public officials can play a key role in investigations of corruption and public accountability. More information about public officials' income, assets, and potential conflict of interests would make it easier to both identify specific individuals of concern (and potentially expose inaccuracies) and to build the public case for stronger ethics laws. Unfortunately, many countries—even countries that require public officials to submit income and asset disclosures—make it too hard for the public to access this information.

**Possibly Relevant Information:** Public financial disclosure forms for U.S. officials. The U.S. government mandates that all officials above GS-15 submit these forms annually. These forms are available to any member of the public by submitting a request. All data is available to members of the public and journalists for free. Personal Financial Disclosure (PFD) parsers are available on GitHub to scrape disclosures into a machine-readable format. Through its "[Trump Town](#)" series, ProPublica series has achieved a workaround of this problem by manually appealing to agencies for their disclosure forms. But no investment has been made to open up these tools to other journalists and the wider public. There are heaps of data out there, but what is missing is international collaboration and usage of non-domestic databases.

**What Tech Tools Might Help?:** One solution is to develop the equivalent of a FOIA machine to access the public financial disclosure forms that all U.S. Government officials above GS-15 rank file (OGE 278 form). Currently these forms are only available by making unique requests to the ethics counsels of individual agencies. An automated, online portal could dramatically streamline this process, as well as enable the creation of a public searchable database (something U.S. lawmakers have considered in the past, but never passed).

A heatmap/twitterbot could automatically scan international databases to compare them with asset disclosures. Are there other tech tools that could be applied to these sources of data to identify anomalies, red flags, etc.? Perhaps by combining data from asset and income declarations with other sources?

## Possible Topic 5: Analyzing Company Ownership and Litigation Data

**The Problem:** Corrupt actors take advantage of corporate secrecy to hide their assets, often using complicated webs of anonymous companies registered in jurisdictions that do not require the true “beneficial owners” of companies to be registered or publicly disclosed.

**Possibly Relevant Information:** The [Offshore Leaks database](#) of the International Consortium of Investigative Journalists contains the metadata of the Offshore Leaks (2013), Panama Papers (2016), Bahamas Leaks (2016), Paradise Papers (2017). It is a treasure trove for anyone interested in investigating corruption. Furthermore, there are at least a few countries (most notably the UK) creating public beneficial ownership registries.

Additionally, Public Access to Court Electronics Records ([PACER](#)) provides online access (for a fee) to U.S. Appellate, District, and Bankruptcy court records and documents nationwide. Although not every lawsuit leads to a conviction, the number of pending cases a company is involved in gives you a sense for their general business behavior – for example in regards to corruption.

**What Tech Tools Might Help?:** Together with specialists from the ICIJ, one could look for new ways of analyzing the data – for example, by finding new graphic visualization methods or by matching it with other pertinent datasets. New tools could also be developed to more efficiently search the datasets of the ICIJ and others.

Connecting Pacer with national procurement databases would enable the creation of a corruption risk heatmap. An alternative approach could be the creation of tools that can scrape and search Pacer and other court filing databases, and connect people or companies involved in legal disputes with those found in other databases. Might there also be a way to scrape Pacer, or other law enforcement databases or news reports, to identify cases (number, type, and location) where anonymous companies were used to facilitate illegal activity (not just, or even mainly, corruption and tax evasion, but also terrorism, drug trafficking, human trafficking, fraud, etc.)?

A tool (possibly a browser-extension similar to the one the Recap-initiative of the Free Law Project is offering) may help to store and make available all downloaded (and paid) Pacer-data at one central place

(like the Court Listener [www.courtlistener.com](http://www.courtlistener.com)). A collaboration with the Free Law Project is highly recommended. Instead of starting from scratch, building on the Free Law Project's efforts makes sense.

## Possible Topic 6: Spotting Red Flags in Public Procurement

**The Problem:** Public procurement (government purchase of goods and services from private vendors) is especially vulnerable to corruption, collusion, fraud, and other sorts of wrongdoing. Although all countries have laws meant to prevent and punish abuses in the procurement process, these laws are not always effective, and sometimes are deliberately left unenforced. We need to bring more transparency to public procurement and raise red flags in the case of dubious deals, including possible connections between political spending (e.g. campaign donations) and the awarding of large-scale public contracts.

**Possibly Relevant Information:** Many countries (and sub-national jurisdictions) have databases of public contracts.

**What Tech Tools Might Help?:** Following the work that [Mihaly Fazekas](#) and others have done, can we develop tech tools that will use public procurement data, perhaps in conjunction with other data sources, to identify suspicious transactions. Can we map correlations between political donations and licenses and contracts made?

## Possible Topic 7: Spotting Red Flags in the Extractive Sector

**The Problem:** The extractive sector presents especially high risks for corruption. The “resource curse” - i.e. the tendency for the presence of abundant raw natural resources to lead to systemically corrupt extractive industries (EI) – is a particularly vexing problem for developing or underdeveloped nations that can ill afford it. Corruption in EI is detrimental to inclusive and diversified economic development, to the rule of law and human rights, and to fair and transparent governance. Foreign corporations are often complicit, and are not adequately held accountable by their home governments.

Understanding and addressing this problem is critical for: 1.) Academic experts in economic development, national and international laws guiding the activities of corporation, political science and governance 2.) Journalists investigating venality and distortion of development through the siphoning off of critical funds, the hollowing out of “commonwealth” enabling dictatorships and bankrolling coercion 3.) Policy makers and activists advocating for fairer and more transparent societies where the resource curse is present, and demanding ethical behavior from the corporations in their own countries

**Possibly Relevant Information:** Though the US is moving backwards on this issue, many countries, including in Europe, are mandating more “publish what you pay” rules for the extractive sector.

**What Tech Tools Might Help?:** Are there ways to combine this information (from the MNC level) with other information (perhaps from the host country) to identify “red flags” for transactions that involved fraud, embezzlement, or other malfeasance? Disclosure of contracts and financial flows in EI, the routes taken by diverted funds, governmental structures potentially enabling EI corruption and benefitting from it, making available data more accessible and easy to understand

## Possible Topic 8: Identifying Individuals and Firms Closely Tied to Senior Government Officials

**The Problem:** Under “know your customer” rules, banks and other financial institutions are supposed to conduct heightened due diligence on so-called “politically exposed persons” (PEPs), such as senior government officials, their family members, and their close friends and associates. Governments do not compile and supply comprehensive PEP lists to financial institutions; rather, those institutions rely on PEP lists supplied by private firms. Those lists, according to many, are both over-inclusive (listing as PEPs people who aren’t actually connected to a senior official but, for example, may have a similar name), and—more worryingly for present purposes—under-inclusive, especially in the case of a kleptocrat’s close associates or cronies. This makes it easier for corrupt actors to evade money laundering controls, by having trusted associates, not identified as PEPs, serve as nominal owners and conduct transactions on their behalf.

**Possibly Relevant Information:** Publicly available data, including natural-language sources (like media accounts), as well as social media may provide some information on the connectedness of various individuals to identified PEPs (such as heads of state and senior ministers). There may also be a great deal of “local knowledge,” for example from domestic civil society organizations and journalists, about business and personal relationships—information that is not currently assimilated into the compilation of existing PEP lists.

**What Tech Tools Might Help?:** There might be ways to enhance the construction of PEP lists (perhaps in collaboration with the firms that already perform this service, perhaps independently) by scraping existing online data sources and/or soliciting input from local human sources, and using this to construct network maps of connectedness to high-risk individuals (e.g. senior government officials from countries with a reputation for high corruption). This information could be used to produce risk scores that could be taken into account when conducting due diligence on potential customers.

Another potential application would be to identify those companies that are owned or controlled by those with close connections to senior government officials, and (building on seminal work by [Ray Fisman](#) and others) assessing how much their economic performance seems to depend on the government official’s health or political security, or the party in control of the government—such information could highlight the role that cronyism may play in some economies and build support for reforms.

## Possible Topic 9: Testing Corporate Registration Agents' Compliance with Verification and Transparency Rules

**The Problem:** Corporate registration agents (CRAs)—those that file the paperwork to create new companies—are supposed to follow certain rules concerning verification of the identity of the party seeking to create the company, and due diligence in case “red flags” indicate possible illegality. However, research by Professors Michael Findley, Daniel Nielson, and Jason Sharman, summarized in their [\*Global Shell Games\*](#) book, found widespread non-compliance. These scholars sent standardized emails to a large sample of CRAs stating an interest in forming a company; some of those emails contained obvious red flags. Yet in a shockingly large percentage of cases, the CRA either indicated a willingness to proceed with the transaction (often without even asking for the legally required identity documentation) or asked for more money to do so. Though this was a one-time scholarly project, it suggests the need for more sustained efforts to identify both individual noncompliance CRAs and the jurisdictions in which such noncompliance is widespread.

**Possibly Relevant Information:** Information on many CRAs can be found online, as these firms advertise their services and often maintain websites. The key information would need to be gathered from them, in a manner similar to that employed by the *Global Shell Games* researchers, but scaled up and sustained over time.

**What Tech Tools Might Help?:** The question is whether there’s a way (both legal and technically feasible) to generalize and automate the process that the *Global Shell Games* authors used to test for CRAs’ compliance with basic requirements: sending a large number of automated emails that appear to be from real people requesting services, with built-in red flags, to identify those not in compliance.



## The Transparathon-Team – Who We Are And How To Reach Out To Us

**Frederik Obermaier**, Investigative Reporter & Initiator of the worldwide Panama Papers/Paradise Papers-investigations, Süddeutsche Zeitung ([Frederik.Obermaier@posteo.de](mailto:Frederik.Obermaier@posteo.de); +49 1781435471)

**Zoe Reiter**, Senior Project Leader and Representative to the U.S. (interim), Transparency International ([zreiter@transparency.org](mailto:zreiter@transparency.org))

**Matthew Stephenson**, Professor, Harvard Law-School ([mstephen@law.harvard.edu](mailto:mstephen@law.harvard.edu))

**David Szakonyi**, Assistant Professor, George Washington University ([davidszakonyi@gmail.com](mailto:davidszakonyi@gmail.com))

**Nikos Passas**, Professor, North-Eastern University ([n.passas@northeastern.edu](mailto:n.passas@northeastern.edu))

**Sherman Teichman**, Senior Fellow, Carr-Center Harvard University ([sherman.teichman@hks.harvard.edu](mailto:sherman.teichman@hks.harvard.edu))

**Raymond Fisman**, Professor, Boston University ([rfisman@bu.edu](mailto:rfisman@bu.edu))