Legal & Ethical Analysis of TransparencyUSA.org

Reidentification in Open Datasets

W231 Tuesdays 6:30 - 8:00PM, Section 2
Connor Brew, Katy Scott, and Samantha Williams

We performed a legal and ethical analysis of TransparencyUSA.org, a website that publishes a database of campaign donations searchable by state, election season, donor, and recipient. The information available in this dataset makes it extremely easy to search for individual donors beyond their campaign donations. A google search for the donor's name, state, and recipient organization typically yields the donor's personal LinkedIn page, Facebook page, and other public-facing information. Digging further into the data by reviewing the recipient organizations of the donor's contributions often quickly reveals a bias toward a political party or campaign topic. Because of the extremely high degree of transparency in this dataset, inconsistent state regulations surrounding its use including secondary uses and the high potential of appropriation and disclosure, we recommend that the Senate Select Committee on Ethics make substantial changes to the regulation surrounding campaign finance and the data surrounding such contributions.

Data Overview

TransparencyUSA.org provides information regarding state and local campaign donations for 12 states: Arizona, California, Florida, Indiana, Michigan, Minnesota, North Carolina, Ohio, Pennsylvania, Texas, Virginia, and Wisconsin. Their mission is to "[P]rovide clear, accurate, easy-to-understand information about the money in state politics."[1] Users of the site can search all state-level candidates, politicians (Governor, State Attorney's General, and legislators at the state House of Representatives and Senate), PACs that are active in that state, and the individual donors that support them.

In reviewing the site, TransparencyUSA does not provide a privacy policy nor a terms and conditions page for users to review or comprehend how the organization might use their data. However, they do address some of these concepts typically contained on these pages in a section located on the FAQs page. According to the information provided on this page, TransparencyUSA advises that "legal uses of our data vary by state", purchase of more tailored datasets are available and that it is best to "consult an attorney licensed in your state for specific information on appropriate data applications."[2] The site will not remove or alter any information provided as certain information is required by FEC but will gladly amend any details once it is changed with the state agency through the candidate, political party, or PAC.

Using the search features for donors and recipient organizations yields a wealth of information. Donor searches will generate lists of donations made by date, amount, and recipient entities and include the donor's name, city, and state. Recipient searches produce donation and expenditure trends, as well as searchable lists of donors and recipients of expenditures.

Current Campaign Finance Laws

For data on federal elections and state data stored at the federal level, the Federal Election Campaign Act affords substantial restrictions surrounding the sharing and use

[1] Transparency USA. (2022, June 16). *About*. https://www.transparencyusa.org/about-us

[2] TransparencyUSA. (2022, June 16). *Frequently Asked Questions*. https://www.transparencyusa.org/faq

of personal information included in FEC data. Additionally, this requires further transparency regarding the personal information of donors who contribute at least $200 per campaign cycle, which must be published openly on the FEC website within 48 hours of receipt. Because of that requirement, most state regulatory bodies publish all of their campaign contribution data to ensure maximum compliance with the Act.

Although the Act ensures protection surrounding the secondary use, appropriation, and disclosure of personal information related to campaign contributions, its protection does not extend to state-managed datasets; the primary sources for TransparencyUSA's data. These state-managed datasets are subject only to the regulations of the state in question, which even after substantial research and investigation are inconsistent and opaque at best.

Privacy Concerns

The state campaign contributions data available on TransparencyUSA.org raises three concerns that compound and exacerbate one another. The first privacy concern that arises is the high degree of transparency. The lack of privacy protection for personal identifiable information exposes individual campaign contributors to risk. Inconsistent legal protections exacerbate this risk because TransparencyUSA.org collects its data from state regulatory bodies[3] rather than the Federal Election Commission meaning that the data hosted on their site is not covered by the FEC's sharing and use restrictions. TransparencyUSA.org data is only protected by the laws and regulations at the state level in which the data originated.

[3] TransparencyUSA. (2022, June 16). *About Us*. https://www.transparencyusa.org/about-us

This leads to our second privacy concern, inconsistent protection varies widely from state to state. As a result, protections around data uses are reduced, creating a complex and unwieldy legal maze that presents an enormous obstacle in understanding what uses of the dataset are legally allowed. For those included in TransparencyUSA's dataset, attempting to discover what protections are available to them is practically impossible.

Finally, the confluence of these earlier risks creates a substantial risk of secondary use abuses, disclosure, and appropriation of data through negligent or malicious targeting of individuals through the data available on TransparencyUSA.org.

Privacy Analysis Using Frameworks

Various frameworks are useful in evaluating potential privacy harms associated with the broad availability of uses for this data set. Solove's taxonomy, Nissenbaum's contextual integrity, Mulligan et al's privacy analytic, and the Belmont Principles together provide structure for this analysis.

Solove's taxonomic framework for the analysis of privacy considers privacy issues through the lens of four categories: Information Collection, Information Processing, Information Dissemination, and Invasion.[4] Solove defines information collection as being either inherently active or passive - interrogation, or surveillance, respectively. The dataset made available by TransparencyUSA.org uses both means to collect data with the majority of it scraped from easily accessible public web interfaces maintained by state regulatory bodies. However, some data is also actively collected and processed

[4] Solove, Daniel J. (2006). A Taxonomy of Privacy. *University of Pennsylvania Law Review,* 154:3 (January 2006), https://scholarship.law.upenn.edu/cgi/viewcontent.cgi?article=1376&context=penn_law_review, p. 477

through a complex bureaucratic request process between TransparencyUSA and the state regulatory body[5]. This process does not quite fit into Solove's definition of interrogation as TransparencyUSA gathers the data from third parties, not the individual donor directly. Therefore, the burden of privacy protection lies both on TransparencyUSA as well as the state regulatory body which collected and disseminated the data in the first place.

Information Processing, the second facet of Solove's taxonomy, is how the data may be connected to an individual and whether or not the individual has provided informed consent about the subsequent processes and applications of that data. Concerning this dataset, the data is directly identifiable to a specific individual. A cursory google search of any single record within the dataset returns a plethora of results for nearly every individual. Of particular importance in this category is Solove's idea of the topics of Secondary Use and Exclusion. Solove points out that, while individuals in the modern age are typically savvy to providing consent to the initial collection and use of their data, they often are not aware of how their data is subsequently used with or without their consent. For example, state regulatory bodies inform individuals that they will collect their personal information for regulatory purposes when a contribution is made and inform them that such information is public record. However, the individual almost certainly did not provide consent for their data to be re-aggregated and re-disseminated in the form of public datasets such as TransparencyUSA. Adding further context by applying the principles of the Belmont Report,[6] "[r]espect for persons requires that

[5] TransparencyUSA. (2022, June 16)., https://www.transparencyusa.org/article/where-did-transparency-usa-get-these-florida-numbers

[6] The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research. April 18, 1979

subjects, to the degree, that they are capable, be given the opportunity to choose what shall or shall not happen to them." TransparencyUSA does not satisfy this principle.

The enormous risk of Exclusion, which Solove defines as "the failure to allow the data subject to know about the data that others have about her and participate in its handling and use,"[7] is enumerated in analytic reports released by TransparencyUSA, as well as the strong risk that other individuals or organizations have the ability and access to aggregate the personal data from TransparencyUSA for potentially malicious purposes.

Of the third category of privacy violations in Solove's taxonomy, many topics are related specifically to the intentional or negligent violation of specific promises made to the individual, as in a breach of confidentiality. These do not apply in this instance, as the state regulatory bodies inform individuals that this data will become public record. However, Solove does identify two topics that are sharply apparent with applicability to TransparencyUSA's dataset: Disclosure, Increased Accessibility, and Appropriation. For disclosure, Solove identifies concerns regarding how the availability of data may cause individuals to be judged differently. For example, an individual who makes a substantial financial donation to a specific PAC or candidate may have their character associated with the beliefs and values of that PAC or candidate, although those values may not be in total alignment.

Consider a scenario where an individual donates a substantial amount of money to a republican candidate because the donor supports the candidate's strong policy on gun control but the donor and the candidate have different views on abortion rights. The

[7] Solove, Daniel J. (2006). A Taxonomy of Privacy. *University of Pennsylvania Law Review,* 154:3 (January 2006), https://scholarship.law.upenn.edu/cgi/viewcontent.cgi?article=1376&context=penn_law_review, p521

donor’s coworkers have seen their financial contributions to this candidate and judge the donor’s character based on all of the candidate’s policies. The donor’s coworkers wrongly judge them based on their political affiliation, which was deciphered from this public information. Increased accessibility is straightforward and does not warrant a deep explanation based on a contextual understanding of the public record. It is unlikely that the common individual understands that their consent to data in the public record also consists of the widespread processing and dissemination of their data via third parties such as TransparencyUSA.

The concern of Appropriation is apparent. Appropriation asks the question: can someone use this data for reasons other than its intended purpose? Donating individuals could be subject to targeting, either for voter suppression or harassment, if antagonists understand their political leanings by viewing their donation history online. This dataset is further concerning when considering job applicants for roles with political aspects, as they may struggle to get a job if they have donated to candidates or committees of another party. Donors may be further targeted not only by campaigns raising funds but also by any retailers interested in segmenting by donation amount or by political leanings. Additionally, when combined with public data related to housing information and other public records, the data could potentially be used for even more malicious purposes.

The comprehensive nature of the information provided in this data set (names, cities, donation amount, and recipient) presents risks of Appropriation related harms beyond those for donating individuals. This data can be easily matched to other publicly available or brokered data sets to identify likely associates of the donors and infer their political preferences and likelihood to donate. This would leave them open to similar risks as those of the individuals who donated, including risks of being targeted for harassment or marketing, and loss of vote confidentiality.

While this data alone does not starkly violate Solove's ideas concerning intrusion, the potential for intrusion is very present. In combination with further research or other open data sources, the risk of intrusion is very high. Solove notes that "Intrusion need not involve spatial incursions: spam, junk mail, junk faxes, and telemarketing are disruptive in a similar way, as they sap people's time and attention and interrupt their activities."[8] The existence of modern automation techniques would easily enable an individual or organization, using this dataset, to spam and solicit individual contributors.

Nissenbaum suggests consideration of the context for when data is collected when determining potential privacy issues.[9] This data is originally provided at the moment of a campaign donation when donors exercise their rights in contributing to the direction in which laws and officials are elected with a record of each transaction stored in a database. There are not many similarities within everyday life, but one could liken it to buying a property in America in that the final sale price is recorded, and the address of the property along with the owner's information is also stored in a public database. In both instances, privacy is lost by the individual who uses their money to exercise fundamental rights: participation in our democracy, and ownership of property. Consent is not requested but assumed, as public records must be kept regarding who owns the land, building, etc., and who is paying taxes on it. Essentially, anything that can be

[8] Solove, Daniel J. (2006). A Taxonomy of Privacy. *University of Pennsylvania Law Review,* 154:3 (January 2006), https://scholarship.law.upenn.edu/cgi/viewcontent.cgi?article=1376&context=penn_law_review, p 550
[9] Nissenbaum, Helen F. (2011). A Contextual Approach to Privacy Online. *Daedalus* 140:4 (Fall 2011), 32-48.

taxed by the government will have a public record for it, including property and campaign donations.

The goal of publishing donor information is transparency which prevents corruption and informs enforcement of limits on certain donation types. This is intended to inform voter choices about the candidates, political parties, or proposed laws.[10] However, this transparency comes at the cost of the donor's privacy. For large donations the benefits are clear: we should know who the people are that are trying to influence our political system. For the smallest donors, the loss of privacy outweighs any social benefits from transparency. According to the Belmont Principle of Beneficence, the benefits to subjects must outweigh the risks of harm.[11] The smaller donor does not necessarily have the same resources as large donors, corporate donors, or political action committees (PACs) in that they may not be able to protect themselves from the potential harms described above. Think of the celebrity that purchases a home, while their address is public record, they can afford to install advanced security systems, hire bodyguards, construct large walls for privacy and take legal measures to ensure their privacy and safety while at their residence. Compare that to a domestic abuse victim who is trying to escape an abuser with limited resources. These individuals are not the same and should not be treated the same.

We need campaign finance laws that address the uniqueness of each donor's contribution. A $500 donation does not have the same influence or weight as a

[10] FEC. (2022, March 28). *Federal Election Commission Fiscal Year 2023 Congressional Budget Justification*, https://Www.Fec.Gov/Resources/Cms-Content/Documents/FEC_FY23_CBJ_March_28_2022.Pdf. Retrieved July 17, 2022, from https://www.fec.gov/resources/cms-content/documents/FEC_FY23_CBJ_March_28_2022.pdf

[11] The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research. April 18, 1979

$500,000 contribution. The TransparencyUSA dataset illustrates how our current system fails to satisfy the Belmont Principle of Justice, which calls for fair treatment and fair distribution of risks and benefits for all.[12] Risks of corruption are better mitigated through donation limits than through transparency which disproportionately harms small donors.

Mulligan et al's privacy analytic[13] provides a useful frame to analyze data presentations like TransparencyUSA for potential risks not just to the individuals and associates of individuals described above, but also for citizens of the United States as a group. Predictive models have the potential to credibly unveil voting histories and preferences for non-donors as well. Concerns regarding these potential harms could make individuals less likely to participate in campaign donations, effectively silencing the will of individuals in favor of larger donors. The subject of this harm is the concept of free and fair elections which underpin our democracy. The protection of confidentiality in elections comes from the State and is upheld by individuals who prefer to keep the information private.

Conclusion and Recommendations

Current practices demonstrated by the TransparencyUSA website violate privacy expectations set by the context of data collection and through a lack of consent. Broadly available campaign donation data has the potential to undermine democracy through various harms to donors, harms which are greater for small value donors versus

[12] The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research. April 18, 1979

[13] Mulligan, Deirdre K., Koopman, Colin and Doty, Nick (2016). Privacy is an essentially contested concept: a multi-dimensional analytic for mapping privacy. *Philosophical Transactions of The Royal Society A: Mathematical Physical and Engineering Sciences*, 374(2083):20160118 (December 2016)

wealthy individual donors or large organizational entities. This can discourage people from supporting campaigns for their candidates of choice and can invite harassment or disinformation from malicious parties which diminish voter turnout. Potential de-anonymization of votes from predictive modeling based on data from donors and their associates could enable greater influence and corruption of our electoral process than campaign funding alone.

In support of individual privacy and free and fair elections, we recommend the following changes:

1. Apply current FEC campaign contribution limits to all elections, including State and Local elections, and to all entities, including SuperPACs. Applying this limitation will reduce the disparity of influence that currently exists, driving the need for transparency to prevent corruption in our process. With less need for transparency, we will have less need to sacrifice privacy.
2. Discontinue publication of identified campaign contributions. While reporting of these contributions to the FEC should continue to ensure compliance with donation limits and other rules, the names, employers, and locations of donors would no longer need to be published.
3. The Federal Election Campaign Act that prevents data on individual donor contributions to federal elections from being used for solicitation or other commercial purposes should be expanded to provide the same protections to

donors for elections at all levels of government in our country.[14] No protections are currently in place for donor data of state or local elections.

4. Penalties associated with violations of the Act should be increased to ensure compliance. Election spending in 2020 was $14.4 billion.[15] In 2020 and 2021, civil penalties issued by the FEC totaled $1,505,878.[16] Given the wealth involved, and the increasing potential to misuse information as data technology continues to evolve, the administrative citations and civil penalties meted out by the FEC are insufficient to deter abuse. We recommend criminal penalties and much steeper fines.

The FEC report cited above publishes metrics on enforcement processes regarding campaign finance violations. Interestingly, *all the personal information of violating individuals is redacted*. Protection of the identities of individuals who have abused their right to participate in elections undermines the transparency which the FEC claims to provide. Data transparency on election donations risks our citizens and weakens our democracy, and in exchange affords questionable protection against corruption. We implore you to consider these reforms.

[14] The Federal Election Campaign Act of 1971, Sec. 311.(a)(4), https://www.govinfo.gov/content/pkg/COMPS-985/pdf/COMPS-985.pdf

[15] OpenSecrets.org. (2021, December 23). *A look back at money-in-politics in 2021*. OpenSecrets News. https://www.opensecrets.org/news/2021/12/a-look-back-at-money-in-politics-2021/

[16] Federal Election Commission, The. Status of Enforcement – Fiscal Year 2022, First Quarter (10/01/21-12/31/21) https://www.fec.gov/resources/cms-content/documents/Status_of_Enforcement_for_the_First_Quarter_of_FY_2022.pdf