NICAR24
March 7–10 • Baltimore

Cracking the courts with data

James V. Grimaldi
The Wall Street Journal
james.grimaldi@wsj.com
@jamesvgrimaldi

William Palin
Free Law Project
bill@free.law
@flooie

Download the Tips Sheet with Links:

https://bit.ly/TipSheetNICAR-CourtData

Getting Started:

- Register for a CourtListener account, if you don’t already have one at www.courtlistener.com
- Get your FREE API key - www.courtlistener.com/profile/api-token

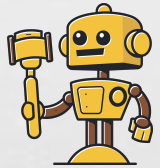
Free Law Project

Free Law Project, a 501(c)(3) non-profit, uses technology, data, and advocacy to democratize access to legal information, regardless of socioeconomic status. We strive to make the law understandable and accessible to all, using advanced technology and data analysis to break down barriers.

Free Law Project

COURT LISTENER

RECAP



What is PACER?

- The **P**ublic **A**ccess to **C**ourt **E**lectronic **R**ecords (PACER) service provides electronic public access to every federal court record.
- About 1.5B records — the largest paywall on the planet.
- Access to case information costs $0.10 per page, capped at $3.00 for most documents.
- Criticisms: out of date, hard to use, expensive, anti-democratic, illegal.

- RECAP is a response that crowdsources documents and makes them freely available.
- Metadata about 350M documents, but only 10M docs.
- Full text searchable.
- Free

COURT LISTENER

What is CourtListener, and how is it helpful?

- CourtListener is a free legal research website containing millions of legal opinions from federal and state courts.
- Fully searchable using real expressions, natural language, and emojis.
- 9M opinions from over a thousand jurisdictions
- Nearly complete SCOTUS record
- More opinions scraped from court websites daily (about 300/day)
- Search and docket alerts for journalists, lawyers, and the public
- API, bulk data, searchable interface

www.courtlistener.com

COURT LISTENER

What is in CourtListener?

- Nearly complete record of US Case Law
  - Case Law Search
  - Citation Look Up
  - Citation Visualizations
- RECAP Archive
- Largest collection of Oral Arguments
- Judges database
- Judicial Financial Disclosures

www.courtlistener.com/help/coverage

RECAP The Law — “Turning PACER Around” since 2009

- Browser extensions for Chrome, Safari, Edge, and Firefox that save you money while you use PACER.
- Everything you purchase gets added to our public archive and gets a permalink.
- Everything anybody else purchased you get for free.
- Use RECAP to save money and contribute to a public archive.
- Over 20k users

👉 https://free.law/recap/ 👈

CourtListener.com - RECAP Archive

- Biggest open collection of federal filings online — 60M cases and 350M items
- Sourced from:
  - Crawling free opinions from PACER
  - RECAP Extension and @recap.email
  - PACER RSS feeds
  - Bulk data gathering services
  - Big Cases Twitter bot & handful of topical case bots
- Full text searchable including scanned docs

CourtListener.com - RECAP Archive

- Excellent for investigative research:
  - Full text search makes background checks more complete
  - Biggest open and searchable collection of federal court filings
- Innumerable historical documents from federal cases:
  - U.S. v. Manafort, U.S. v. Gates, U.S. v. Michael Cohen
  - The “Citizenship on the Census” case
  - “DNC v. Russian Trolls” case
  - Trump Emoluments
- Alerts for dockets to know when new filings are available
- Services available for bulk research without having interns click thousands of PACER documents

RECAP Demo

Recorded Demo: https://youtu.be/NdfuLlcL1CU

Building the Judicial Financial Disclosures Database

What Started it

- 1978 Ethics in Government Act
- 2017 Removal of the per disclosure fee

Paper Files

FLP began requesting the files in bulk starting in 2017.

Machine Learning

874 disclosure records added going back to 1987 from Congressional archives

Court Websites

Scraping of online sources provided some disclosures from 2003 to 2010.

Judicial Financial Disclosures Dataset

https://www.courtlistener.com/financial-disclosures/

Federal Judicial Financial Disclosures

Justices' and judges' financial disclosures must be provided to "any person" who requests them.

– the Ethics in Government Act.

AO 10
Rev. 1/2021

FINANCIAL DISCLOSURE REPORT
FOR CALENDAR YEAR 2020

Report Required by the Ethics in Government Act of 1978 (5 U.S.C. app. §§ 101-111)

1. Person Reporting (last name, first, middle initial)	2. Court or Organization	3. Date of Report
Gilstrap, James R.	US District Court, Eastern District of Texas	12/07/2021
4. Title (Article III judges indicate active or senior status; magistrate judges indicate full- or part-time)	**5a. Report Type (check appropriate type)**	**6. Reporting Period**
United States District Judge (Active)	☐ Nomination Date ☐ Initial ☑ Annual ☐ Final 5b. ☑ Amended Report	01/01/2020 to 12/31/2020
7. Chambers or Office Address		
100 East Houston Street Marshall, TX 75670		

***IMPORTANT NOTES:** The instructions accompanying this form must be followed. Complete all parts, checking the NONE box for each part where you have no reportable information.*

I. POSITIONS. *(Reporting individual only; see pp. 9-13 of filing instructions)*

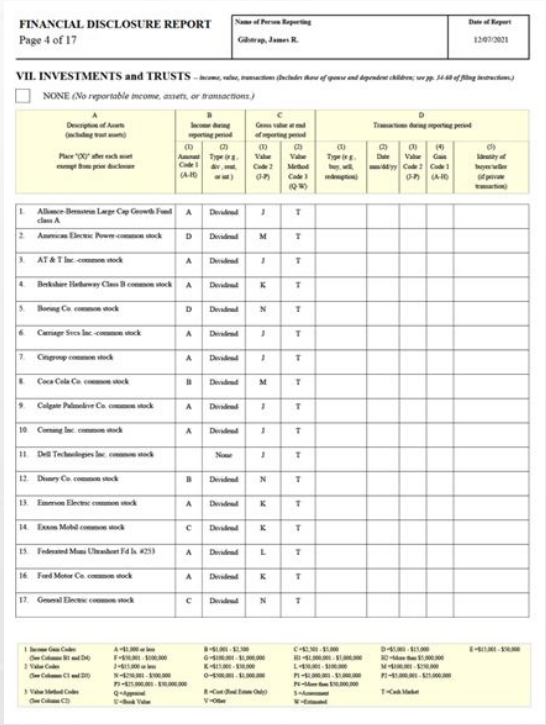
☐ NONE *(No reportable positions.)*

	POSITION	NAME OF ORGANIZATION/ENTITY
1.	Trustee	Tommy B. & Lucille Jackson Slaughter Foundation No. 2
2.	Trustee	T. Whitfield Davidson Foundation
3.		
4.		
5.		

II. AGREEMENTS. *(Reporting individual only; see pp. 14-16 of filing instructions)*

☑ NONE *(No reportable agreements.)*

	DATE	PARTIES AND TERMS
1.		
2.		
3.		



FINANCIAL DISCLOSURE REPORT
Page 4 of 17

Name of Person Reporting	Date of Report
Gilstrap, James R.	12/07/2021

VII. INVESTMENTS and TRUSTS – *Income, value, transactions (Includes those of spouse and dependent children; see pp. 34-60 of filing instructions.)*

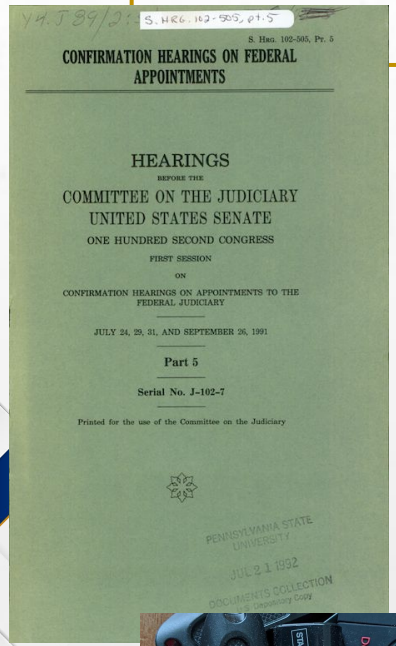
☐ NONE *(No reportable income, assets, or transactions.)*

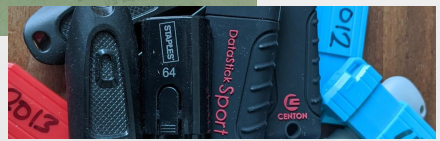
A Description of Assets (including trust assets) Place "(X)" after each asset exempt from prior disclosure	B Income during reporting period (1) Amount Code 1 (A-H)	B (2) Type (e.g., div., rent, or int.)	C Gross value at end of reporting period (1) Value Code 2 (J-P)	C (2) Value Method Code 3 (Q-W)	D Transactions during reporting period (1) Type (e.g., buy, sell, redemption)	(2) Date mm/dd/yy	(3) Value Code 2 (J-P)	(4) Gain Code 1 (A-H)	(5) Identity of buyer/seller (if private transaction)
1. Alliance-Bernstein Large Cap Growth Fund class A	A	Dividend	J	T					
2. American Electric Power-common stock	D	Dividend	M	T					
3. AT & T Inc.-common stock	A	Dividend	J	T					
4. Berkshire Hathaway Class B common stock	A	Dividend	K	T					
5. Boeing Co. common stock	D	Dividend	N	T					
6. Carriage Svcs Inc.-common stock	A	Dividend	J	T					
7. Citigroup common stock	A	Dividend	J	T					
8. Coca Cola Co. common stock	B	Dividend	M	T					
9. Colgate Palmolive Co. common stock	A	Dividend	J	T					
10. Corning Inc. common stock	A	Dividend	J	T					
11. Dell Technologies Inc. common stock		None	J	T					
12. Disney Co. common stock	B	Dividend	N	T					
13. Emerson Electric common stock	A	Dividend	K	T					
14. Exxon Mobil common stock	C	Dividend	K	T					
15. Federated Muni Ultrashort Fd Is. #253	A	Dividend	L	T					
16. Ford Motor Co. common stock	A	Dividend	K	T					
17. General Electric common stock	C	Dividend	N	T					

1 Income Gain Codes (See Columns B1 and D4)	A =$1,000 or less F =$50,001 - $100,000 J =$15,000 or less N =$250,001 - $500,000 P3 =$5,000,001 - $50,000,000	B =$1,001 - $2,500 G =$100,001 - $1,000,000 K =$15,001 - $50,000 O =$500,001 - $1,000,000	C =$2,501 - $5,000 H1 =$1,000,001 - $5,000,000 L =$50,001 - $100,000 P1 =$1,000,001 - $5,000,000 P4 =More than $50,000,000	D =$5,001 - $15,000 H2 =More than $5,000,000 M =$100,001 - $250,000 P2 =$5,000,001 - $25,000,000	E =$15,001 - $50,000
2 Value Codes (See Columns C1 and D3)			S =Assessment	T =Cash Market	
3 Value Method Codes (See Column C1)	Q =Appraisal U =Book Value	R =Cost (Real Estate Only) V =Other	W =Estimated		

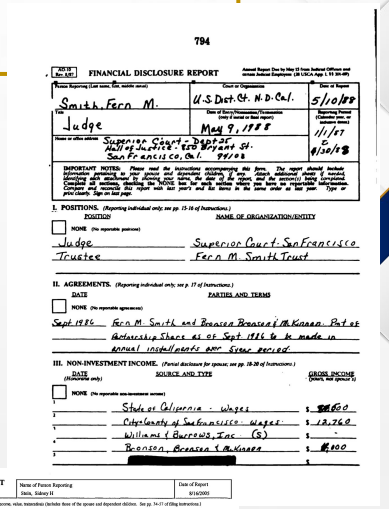
Form AO 10

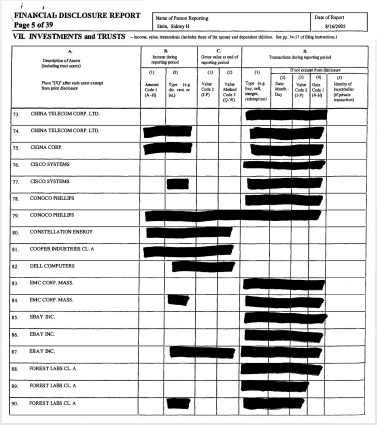
What Free Law Project Collected





- Really long TIFFs - https://bit.ly/longTIFF
- Hundreds of TIFFs for one disclosure - https://bit.ly/100sTIFFS
- Files on jump drives
- 250K pages of documents from Senate hearing records
- Forms with redactions, corrections, misspellings, old companies





Enhancing the Data:

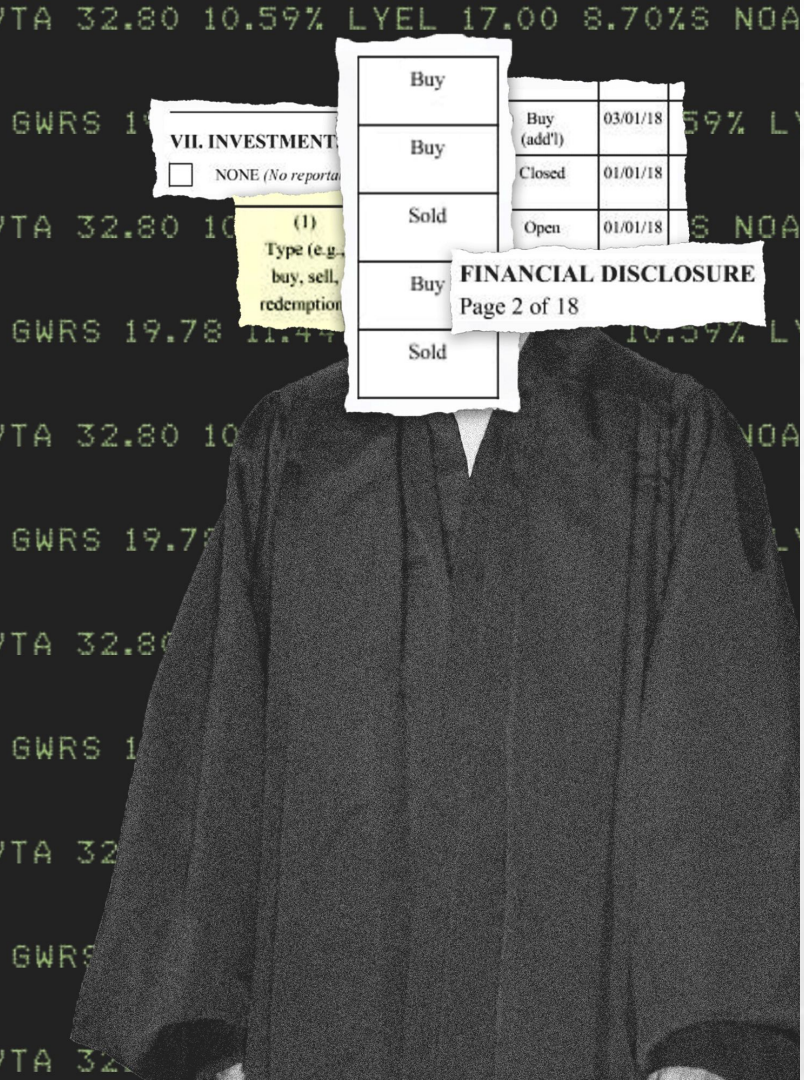
- Free Law Project developed a parsing system for the disclosures.
  - Takes the disclosure, whether it's a single-page TIFF, multi-page TIFF or PDF, scanned document or text, and converts it into machine-readable JSON data.
- A machine learning model to identify pages in the Senate hearing records that looked like financial disclosure documents.
  - Extracted 843 Financial Disclosure from over 250k pages of documents.
- Disclosures were converted into good PDF files and ingested the data into the database
- All data available via CourtListener APIs and website
- Connection of Financial Disclosures dataset with FLP's Judges database

What's in the Dataset? *(So Far)*

32,336 disclosures documenting:

- **1,901,720** investments
- **37,050** positions held outside of the judiciary
- **10,007** legal agreements
- **15,302** sources of non-investment income
- **20,175** sources of spousal income
- **33,472** reimbursements
- **2,028** gifts to judges and judicial employees
- **18,775** debts

Coverage Page: www.courtlistener.com/help/coverage/financial-disclosures/



Investigating Judges and the Courts

WSJ

Hidden Interests

Project with James Grimaldi, Joe Palazzolo, & Coulter Jones

https://on.wsj.com/3tRxpla

Project Idea

Federal law prohibits judges and justices from hearing cases in which they or their families have a “legal or equitable interest, however small” in order to promote confidence in the judiciary.

– 28 U.S. Code § 455 (d)(4)

Tip: Compare data sets to find stories

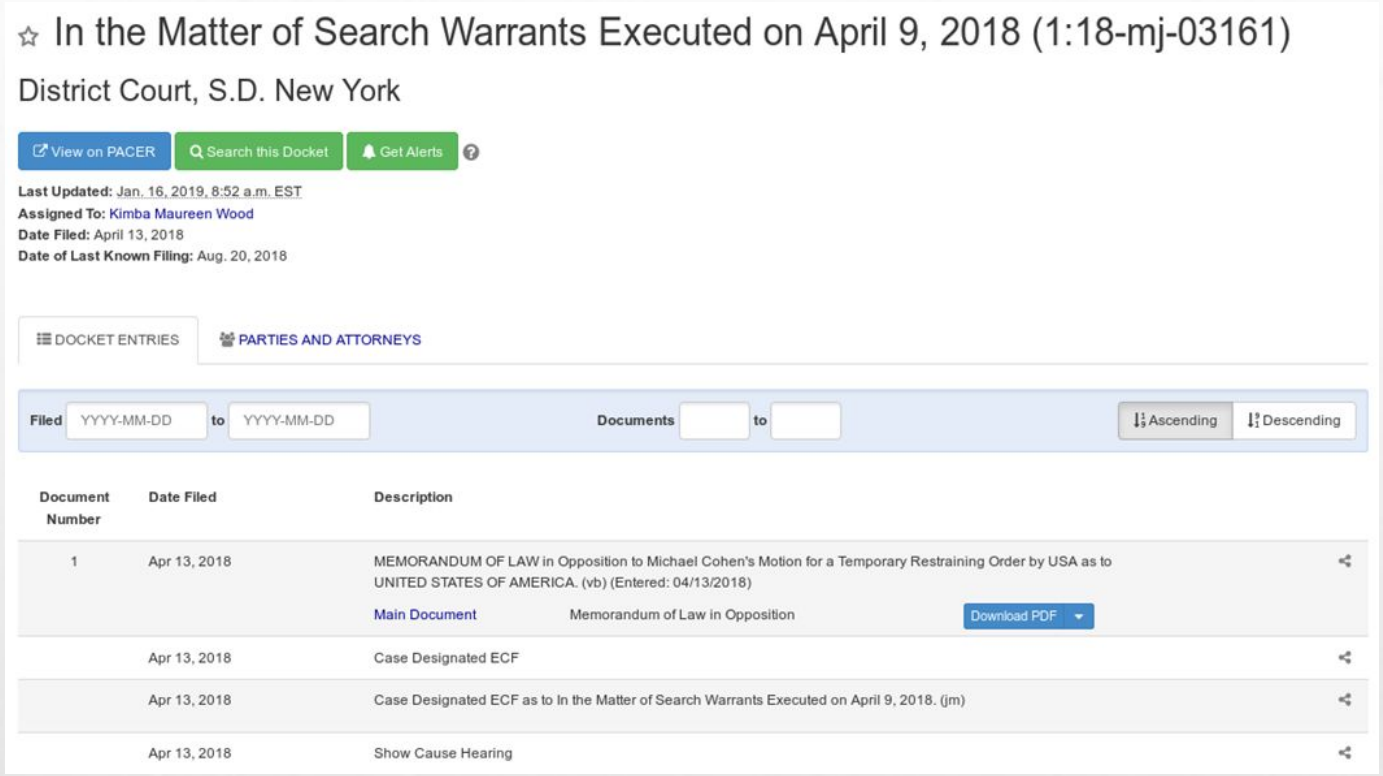
Judges' Dockets

- WSJ and Free Law Project scraped district court websites to obtain the dockets of all judges
- The Journal compared judges' financial disclosure forms against their dockets and created long lists of potential conflicts
- James, Joe and Coulter went through thousands of potential conflicts using PACER

Tip: Whenever covering a federal suit, look up the judge's disclosure

www.courtlistener.com/person

RECAP Dockets



☆ In the Matter of Search Warrants Executed on April 9, 2018 (1:18-mj-03161)

District Court, S.D. New York

View on PACER | Search this Docket | Get Alerts

Last Updated: Jan. 16, 2019, 8:52 a.m. EST
Assigned To: Kimba Maureen Wood
Date Filed: April 13, 2018
Date of Last Known Filing: Aug. 20, 2018

DOCKET ENTRIES | PARTIES AND ATTORNEYS

Filed YYYY-MM-DD to YYYY-MM-DD Documents to Ascending | Descending

Document Number	Date Filed	Description
1	Apr 13, 2018	MEMORANDUM OF LAW in Opposition to Michael Cohen's Motion for a Temporary Restraining Order by USA as to UNITED STATES OF AMERICA. (vb) (Entered: 04/13/2018) Main Document Memorandum of Law in Opposition Download PDF
	Apr 13, 2018	Case Designated ECF
	Apr 13, 2018	Case Designated ECF as to In the Matter of Search Warrants Executed on April 9, 2018. (jm)
	Apr 13, 2018	Show Cause Hearing

Dossiers

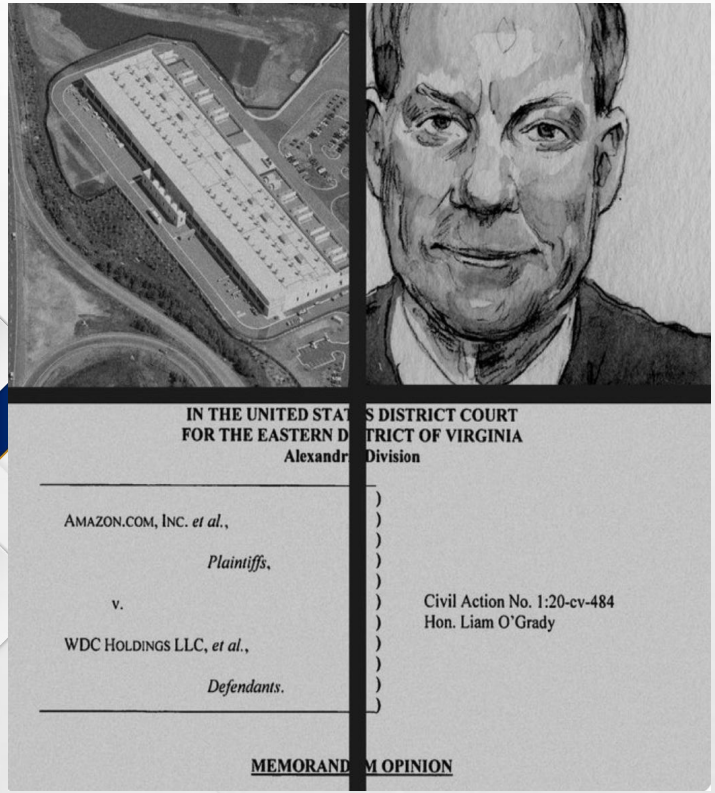
We prepared detailed lists of each recusal violation we found and then contacted the judges

ExxonMobil Oil Corporation v. TIG Insurance Company

Judge Ramos's 2016 and 2017 financial disclosure forms show that a brokerage account held $15,001 to $50,000 of Exxon Mobil Corp. common stock in those years (lines 92 and 82). His 2018 disclosure form shows that the same brokerage account held $15,000 or less of Exxon stock (line 78).

Judge Ramos was assigned to ExxonMobil Oil Corp. v. TIG Insurance Co. (1:16-cv-09527) on 12-12-2016. He granted Exxon's motion to compel arbitration on 2-2-2017 and confirmed the arbitration award on 5-18-2020. The case is currently on appeal.

Findings



- 152 federal judges around the nation have violated U.S. law and judicial ethics by overseeing 1,076 court cases involving companies in which they or their family owned stock.
- 61 judges didn't just own stocks of companies that were litigants. Accounts held by the judges or their families ***traded*** shares as suits were progressing.

on.wsj.com/3zU8a5Q

In New York, U.S. District Judge Edgardo Ramos handled a suit between an Exxon Mobil Corp. unit and TIG Insurance Co. over a pollution claim while owning between $15,001 and $50,000 of Exxon stock. He accepted an arbitration panel's opinion that TIG should pay Exxon $25 million and added $8 million of interest to the tab.

Tip: Tap expertise of nonpartisan legal experts

Results

- After Journal stories, judges in 883 cases have notified courts that they improperly the cases that now are eligible to be reopened
- The U.S. Senate and House immediately responded with a bill to speed up online posting of financial disclosures.

President Biden Signs Judicial Ethics Law Bolstering Stock Disclosures

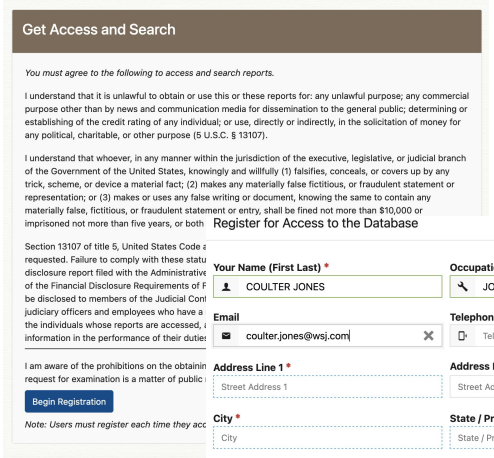
Supreme Court justices, federal judges now required to promptly report their finances; legislation comes after WSJ investigation

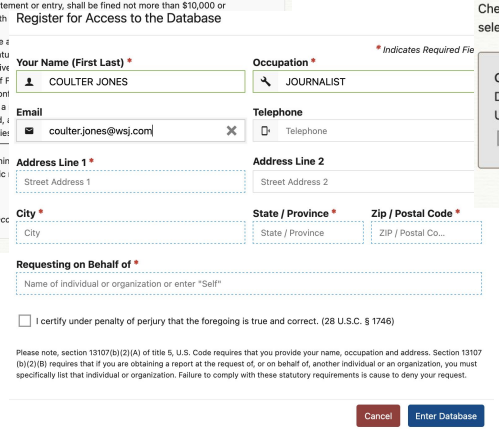
on.wsj.com/3bmQ3Lz

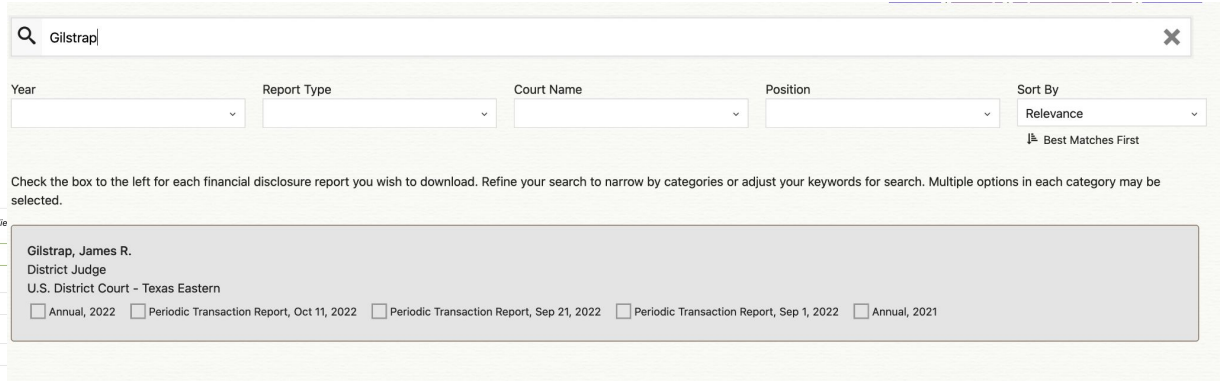
Tip: Contact congressional judiciary committees for expertise

Flaws in the Data

Newer forms: pub.jefs.uscourts.gov/



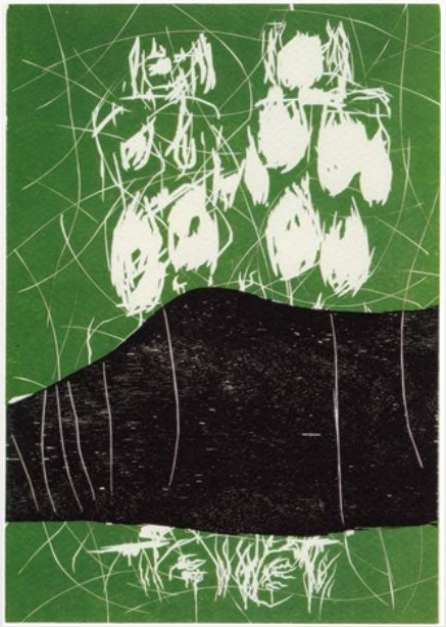




What Else Can be Done With The Data?

Inside the Data: Art Collections

Millions of dollars in art investments



Georg Baselitz, *Der Berg,* 1991

George Condo, *Conversations*, 2011

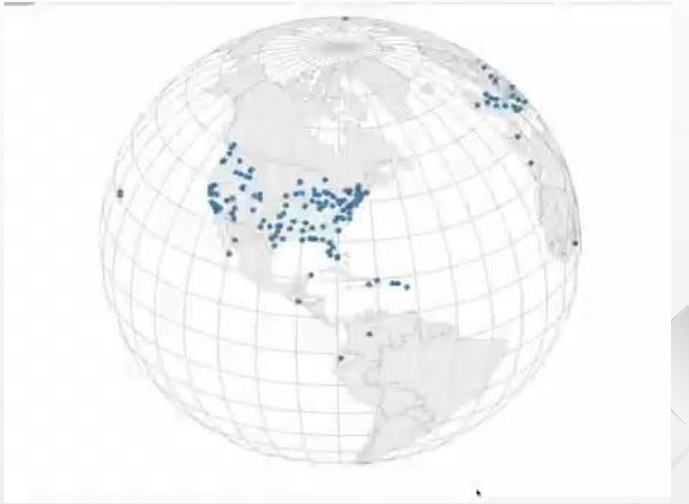
George Condo, *Blue Expanding Figures*, 2006

Examples of Artwork Owned by Judge Helene N. White, U. S. Court of Appeals for the Sixth Circuit

API CALL -> /investments/?description__icontains=Artwork:

Inside the Data: World Travelers

Travel Reimbursements to over 600 locations across the globe.



FINANCIAL DISCLOSURE REPORT
Page 3 of 6

Name of Person Reporting	Date of Report
Davis, Michael J	04/18/2008

V. GIFTS. *(Includes those to spouse and dependent children; see pp. 28-31 of filing instructions.)*

☐ NONE *(No reportable gifts.)*

	SOURCE	DESCRIPTION	VALUE
1.	Minneapolis Club	Honorary Membership	$ 4,500.00
2.	Kingdom of Saudi Arabia Government	2 brief cases, Figs, crystal obelisk	$ 150.00
3.			

13.	Japanese Government	Nov. 17-22, 2007	Tokyo, Japan	Educational conference	travel and lodging
14.	Judicial Yuan of Taiwanese Government	Nov. 23-27, 2007	Taipei, China	Educational conference	travel and lodging

API CALL -> /reimbursements/?source__icontains=Saudi OR /reimbursements/?source__icontains=figs

Inside the Data: Investments of the 1%

9,653 investments greater than $1,000,001 *

- 8,588 investments between $1,000,001 and $5,000,000
- 613 investments between $5,000,001 and $25,000,000
- 64 investments between $25,000,001 and $50,000,000*
- 118 investments above $50,000,001

* This does not account for duplicate investments over multiple years

API CALL -> /investments/?gross_value_code=P3 OR /investments/?gross_value_code=P4

Redactions:

https://storage.courtlistener.com/us/federal/judicial/financial-disclosures/8587/matthew-joseph-kacsmaryk-disclosure.2016.pdf

https://storage.courtlistener.com/us/federal/judicial/financial-disclosures/8587/matthew-joseph-kacsmaryk-disclosure.2021_1.pdf

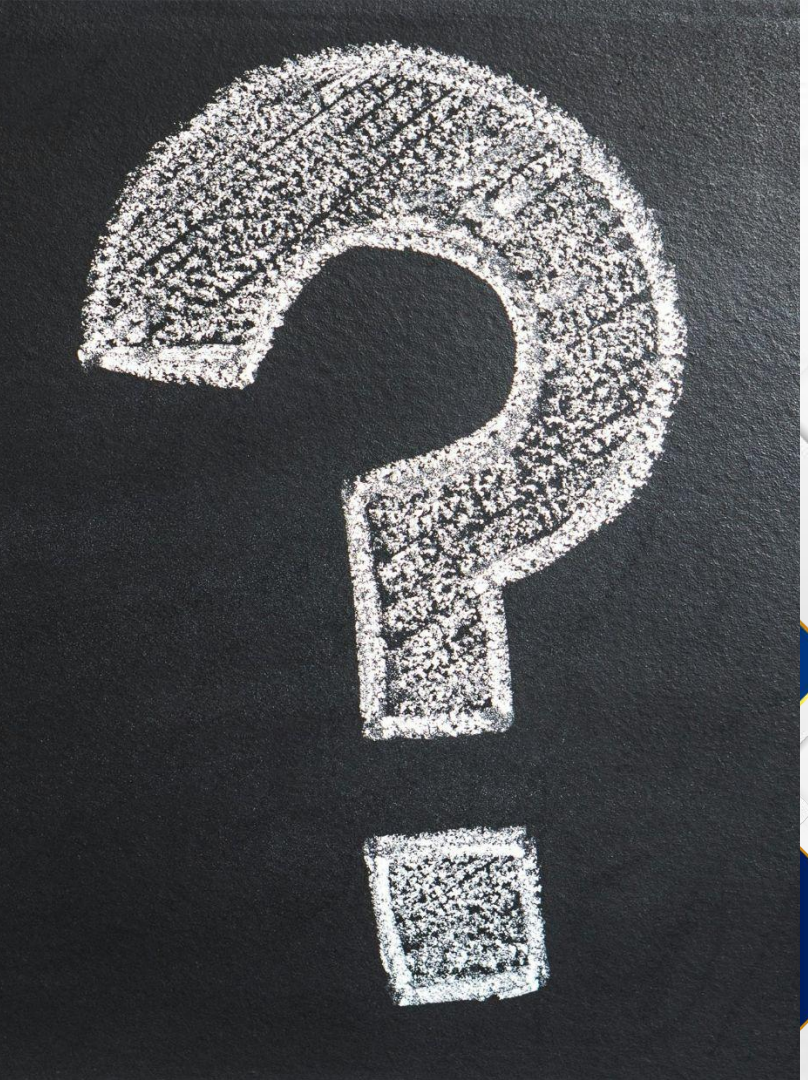
10.	College Invest 529 Aggressive Growth Allocation	A	Distribution	J	T				
11.	Publix Supermarkets, Inc., Common Stock (Private)	G	Distribution	P1	U				
12.	Wells Fargo Bank Accounts	A	Interest	K	T				
13.									

13.	CollegeInvest Age-Based Aggressive Growth	B	Dividend	K	T	
14.	[redacted] Common Stock	G	Dividend	P2	T	
15.						

Hands On

“Information is power. But like all power, there are those who want to keep it for themselves.”

—AARON SWARTZ



Questions?

Thank You!

James V. Grimaldi
The Wall Street Journal
james.grimaldi@wsj.com
@jamesvgrimaldi

William Palin
Free Law Project
bill@free.law
@flooie

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, infographics from **Freepik**, and images from **Pexels**.

Converting to PDFs:

- Disclosures all converted into good PDFs using Doctor. Free Law Project's Open Source microservice for converting and extracting documents and audio files at scale.

https://free.law/projects/doctor

Disclosure Extractor

- Bill - add short description of extractor here

Link to Extractor

Query into Dockets

- Using CourtListener Search - https://www.courtlistener.com/recap/
- Using Python and how to make modifications - NEED LINK