# Hacks, Leaks, and Revelations

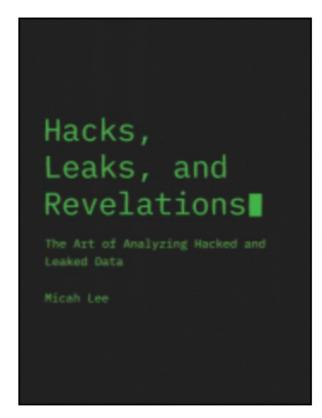
A hands-on guide to analyzing hacked and leaked data

#### **Micah Lee**

Micah Lee is The Intercept's Director of Information Security. He is a computer security engineer and an open-source software developer who writes about technical topics, leaked datasets, and the far right. He develops security and privacy tools such as OnionShare, Dangerzone, and Semiphemeral. He's writing a book to teach journalists and researchers how to analyze leaked datasets called *Hacks, Leaks, and Revelations*, to be released later in 2023.

Before joining The Intercept, he worked as a staff technologist at the Electronic Frontier Foundation, where he explained how technologies work to journalists and lawyers, and worked to encrypt the web. He is a founder and former board member of the Freedom of the Press Foundation, a member of the Distributed Denial of Secrets advisory board, and a Tor Project core contributor.

Mastodon: @micahflee@infosec.social | micah@micahflee.com



#### Hacks, Leaks, and Revelations

The Art of Analyzing Hacked and Leaked Data by Micah Lee

July 2023, 352 pp. ISBN-13: 9781718503120

- Print Book (PREORDER) and EARLY ACCESS Ebook, \$49.99
- EARLY ACCESS Ebook, \$39.99

Pre-Order

Use coupon code PREORDER to get 25% off!

https://nostarch.com/hacks-leaks-and-revelations



# Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)

# The Internet is Overflowing in Hacked and Leaked Datasets

Just in the first two and a half months of 2023...

### Release: Cellebrite (1.7 TB) and MSAB (103 GB)

An anonymous whistleblower has leaked Cellebrite's phone forensics software and its documentation. Co-published with Enlace Hacktivista

#### **Distributed Denial of Secrets**

Jan 13



We are publishing phone forensics software and documentation from the Israeli company Cellebrite and from its Swedish competitor, MSAB. These companies sell their tools to police and governments around the world. Cellebrite and MSAB's tools are typically used to collect information from smartphones.

This data was first published by Enlace Hacktivista, who wrote on their front page:

Jan 13, 2023: An anonymous whistleblower sent us phone forensics software and documentation from Cellebrite and MSAB. These companies sell to police and

## Limited distribution: Fundação Nacional de Artes (302 GB)

Emails from 2011 to 2022 of Brazil's national arts foundation, including the tumultuous Bolsonaro years. Funarte develops policies to promote the arts.

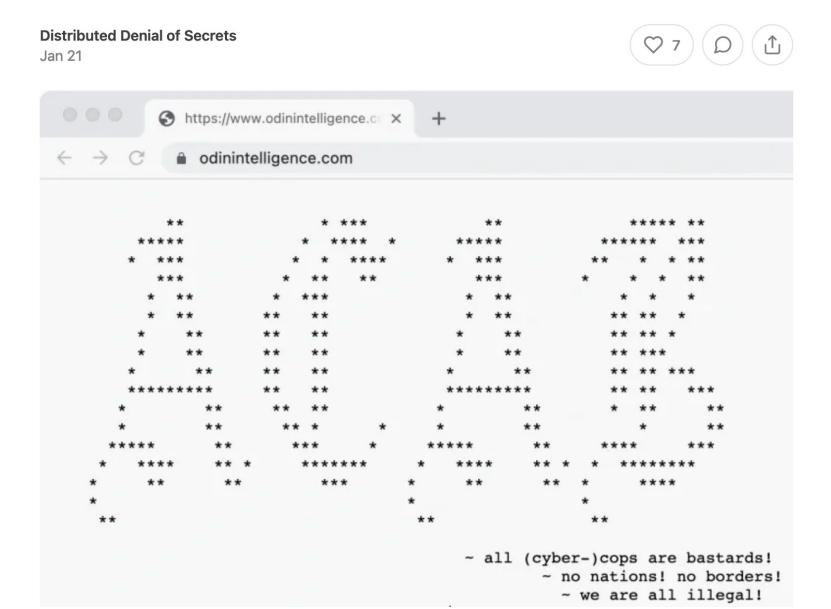
**Distributed Denial of Secrets**Jan 14





#### **Limited distribution: ODIN Intelligence (19 GB)**

A contractor that works for police agencies had its website defaced as reported by Techcrunch



#### **Limited distribution: No Fly list (90 MB)**

A 2019 version of the United States' No Fly List. The TSA's terrorist screening list has more than 1.5 million names including 3000 minors

**Distributed Denial of Secrets**Jan 27





## Limited distribution: Russian Censor Files (335 GB)

Emails and files from the General Radio Frequency Center of the censorship agency Roskomnadzor, via the Belarusian Cyber-partisans

#### **Distributed Denial of Secrets**

Feb 14



Almost a year after the Russian invasion of Ukraine, we have indexed a new leak from Russia's censorship agency Roskomnadzor. These documents and emails come from the internal network of the General Radio Frequency Center (GRFC) subdivision of Roskomnadzor. GRFC monitors social media and other open sources in order to implement the Kremlin's media policies and provide data for decision makers. The GRFC surveils Putin's opponents, writes reports to the FSB & other agencies, and blocks access to independent media sources on the internet.

This is a new dataset, and is different from the 820 GB leak we published in March 2022 from the Republic of Bashkortostan's office of Roskomnadzor.

#### ddosecrets 2-1524-22.pdf

FIREBRE

BINETIA POCCERÉNCIÓ O EXPAIDE

I ma 2017 (n.)

Topos homos o compos

Separas passima de es reputa homos o compos

Separas passima de est reputa homos o compos

Separas passima de est reputa homos o compos

Separas passima de est reputa de est reputa de la 1515/2021 en

Topos de est reputa de est reputa de est reputa de la 1515/2021 en

Topos de est reputa de est reputa de est reputa de la 1515/2021 en

Topos de est reputado de est reputado en anome o a chor est es esperamento

TOPOS de est reputado est reputado en anome o a chor est est reputado en

TOPOS de est reputado en esta en esta de estado en esta de estado en esta de esta



Практивнення метан по допровожен Консоно Т.А. в пестователя срадине месания Практивнення интегнал править месания в песто по техня и места дополня и практивнення менения перавить месания в песто по техня и места дополня от править в менения песто по техня песто песто песто песто песто песто перативнення менения песто песто песто песто песто песто песто песто песто перативнения и песто пе

Супретнений вакти, призвы мистиме своиродинения, принимания инвидиа, принимания обращения в при и напримення в нашту принимания в нашту принимания в нашту при на при на

Останов с на 2 ст. 23 маня об переняльных дивнего Федеральная страба по надхору в офер с меж. запримененных об переняльных дивнего Федеральная страба по подосночными орган по законте цена субъения вересавальных дивне, запраме обращего и уба в пожими заплеменных в дишего две субъения вересавальных дивнях, в том чесле в закоту грав вопоражденного куга на де пределжаться витером субъения вероматира.

В опентентия от с. 3 90 г/о произвания динно (даже - Вин о Процимальная динно даже - Вин о Процимальная динно (даже - Вин о Процимальная динно даже - Вин о Процимальная динно даже - Вин о В

 сойть.
 Интириет-сайт предоставляет домути в файдам, одприжания интеграстторударственные и кумпаравляные служищее, отприжана ризнас компаний и собъектов персоказаваем данных в объект филосия, ком, счетства, даля росказывае, дагод дойстронных раст.

Вмест є тем, на визернет – ресурсі https://ddescrets.com/ отсутствуєт виформация, кіньть, забо вика вильчне согласня граждан на обрабочку и рактространенне як персональных кіньть, забо вика маненные основання за обрабочку указаннями предоставлями динами. Динами обстоятельства подпоряжданням интегнациями получаннями.

в обеспекате свем урбеница предостава, основника предостава, основника образования предостава, пофразования изголенский изистава конценствому образования предостава, информация урожно, урбеница образования услуги, урбеница образования услуги, урбеница образования услуги, урбеница образования от 27 нарад 2006 году урбеница образования от 27 нарад 2006 году урбеница образования образования от 27 нарад 2006 году урбеница образования обр



#### РЕШЕНИЕ ИМЕНЕМ РОСССИЙСКОЙ ФЕДЕРАЦИИ

06 июня 2022 года

город Москва

Таганский районный суд города Москвы в составе председательствующего судьи Синельниковой О.В. при секретаре Адхамжанове А.А.,

рассмотрев в открытом судебном заседании гражданское дело № 2-1524/2022 по исковому заявлению Федеральной службы по надзору в сфере связи, информационных технологий и массовых коммуникаций к Internet Domain Service BS Corp в защиту прав неопределенного круга лиц,

#### УСТАНОВИЛ:

Истец обратился в суд с иском к ответчику Internet Domain Service BS Corp о признании деятельности интернет – ресурсов незаконной, признании информации запрещенной, указывая в обоснование иска на то, что Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций по защите прав субъектов персональных данных в ходе мониторинга информационно-телекоммуникационной сети «Интернет» было выявлено нарушение интернет - сайтом по адресу https://ddosecrets.com/ прав и законных интересов граждан как субъектов персональных при в предональных при в при в при в предональных при в предональных при в при в при в при в при в предональных при в при

#### Limited distribution: Oakland City Hall (11.7 GB)

Emails and files from the PLAY ransomware attack on Oakland City Hall, a large city in California with a long history of police abuses.

#### **Distributed Denial of Secrets**

Mar 8



DDoSecrets is making available a series of internal emails and files from Oakland City Hall, a mid-sized city in the San Francisco Bay area with a long history of human rights violations. The files contain information about police assignments for active duty employees, lawsuit settlement agreements, social security numbers, and more.

The files were released after a ransomware negotiation that lasted about a month. The 11.7 GB sample released by the PLAY ransomware hackers contains information about ongoing litigation against the city, wire transfer records, bond sale information, and contracting data. DDoSecrets reviewed some of the files and has confirmed they contain officer disciplinary records including for supervisors who failed to report misconduct. We are also able to confirm the files include information on misconduct allegations against high-ranking police officers, and documents about internal affairs investigations.

# Source Protection, Digital Security, and Acquiring Datasets

### Protecting Your Sources and Yourself

- Protecting sources by minimizing the data trail
- Securely storing datasets
- Verifying that data is authentic using OSINT
- Deciding what to redact
- Using a password manager, using disk encryption, and opening sketchy documents using Dangerzone
- Talking to sources using Signal

### Acquiring Datasets

- Public datasets from DDoSecrets
- Downloading public datasets using BitTorrent
- Various ways to securely get datasets directly from sources

### The Command Line Interface

#### Filtering for Documents Mentioning Antifa

You'll start by grepping our list of filenames for files that include *antifa* in their names. (Later on you'll search the content of files, but searching filenames is quicker and provides a good starting point.) From the *BlueLeaks-extracted* folder, search the *BlueLeaks-filenames.txt* file that you created in Homework 4-2 by running:

```
cat ../BlueLeaks-filenames.txt | grep antifa
```

This command pipes the output of cat\_\_\_./BlueLeaks-filenames.txt (which is a list of a million filenames) as input into grep antifa. Essentially, it filters that huge list of filenames to only show you ones that include the word *antifa*.

That command returns no results. Since the grep command is case-sensitive, try again using the -i, or --ignore-case, argument:

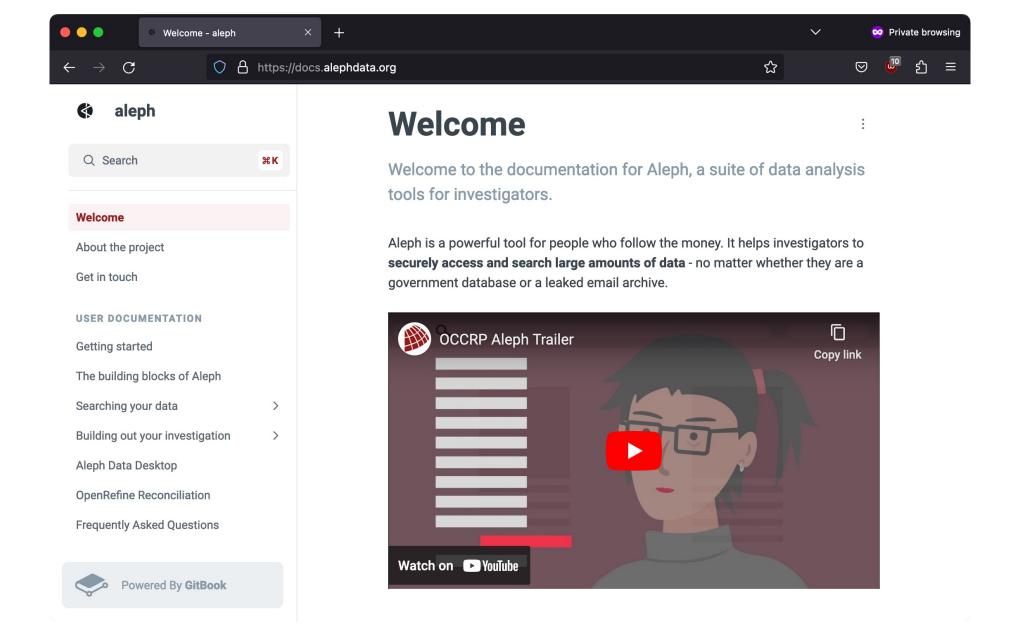
```
cat ../BlueLeaks-filenames.txt | grep -i antifa
```

When I run this command on my macOS computer, this is the output:

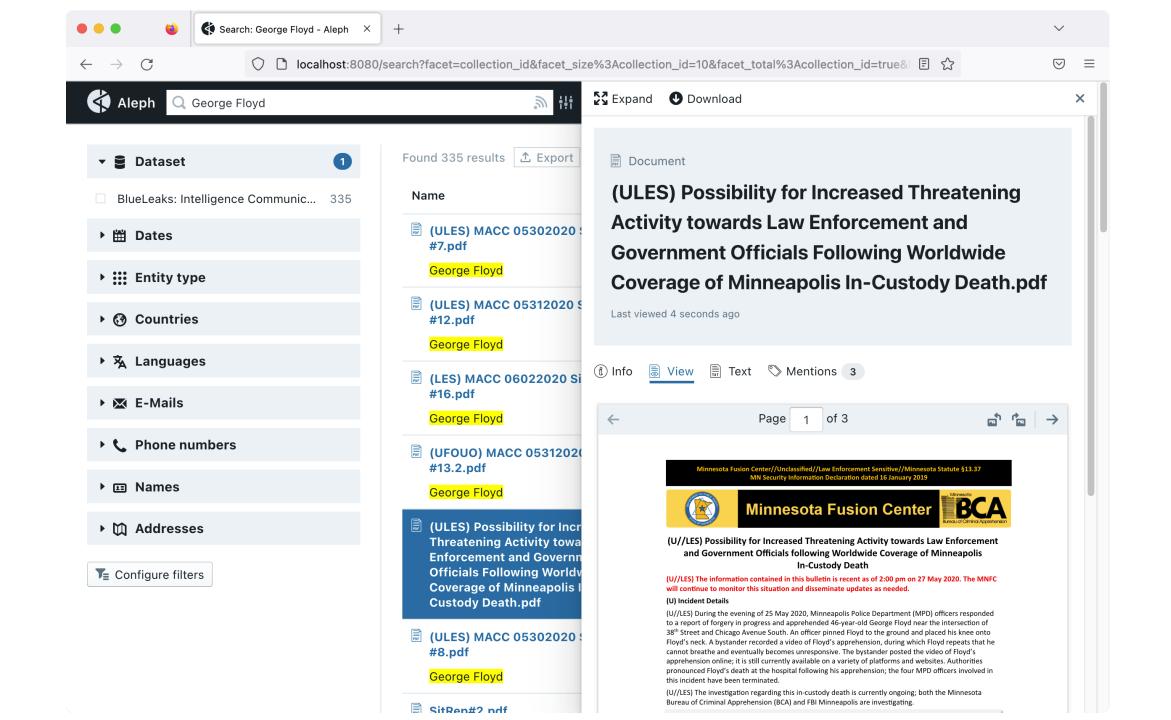
```
./ociac/files/EBAT1/U-FOUO_CFIX_OCIAC_JRA_DVE Use of Social Media_ANTIFA_ANTI-ANTIFA
MOVEMENTS.pdf
./arictexas/files/DDF/ARIC-LES - Situational Awareness - Antifa Activity.pdf
./arictexas/files/DDF/SWTFC-LES - Situational Awareness - ANTIFA Event Notification.pdf
./arictexas/files/DPI/ARIC-LES - Situational Awareness - Antifa Activity.png
./arictexas/files/DPI/SWTFC-LES - Situational Awareness - ANTIFA Event Notification.png
./dediac/files/DDF/ANTIFA - Fighting in the Streets.pdf
./dediac/files/DDF/ANTIFA Sub Groups and Indicators - LES.pdf
./dediac/files/DDF/
FBI_PH_SIR_Tactics_and_Targets_Identified_for_4_November_2017_ANTIFA_Rally_in_Philadelphia_PA-
2.pdf
./dediac/files/EBAT1/ANTIFA - Fighting in the Streets.pdf
./dediac/files/EBAT1/ANTIFA Sub Groups and Indicators - LES.pdf
./dediac/files/DPI/ANTIFA - Fighting in the Streets.png
./dediac/files/DPI/
FBI_PH_SIR_Tactics_and_Targets_Identified_for_4_November_2017_ANTIFA_Rally_in_Philadelphia_PA-
2.png
```

This command returns 12 results, all files that have the term *antifa* in their filenames. The grep command might highlight your search terms in each line of output by coloring them differently; I've highlighted them in this output by boldfacing them. Try reading some of the documents in this list.

# Making Datasets Searchable with Aleph



https://docs.alephdata.org/ | https://data.occrp.org/



# German authorities seize 'BlueLeaks' server that hosted data on US cops

BlueLeaks portal is now down. The website hosted 269 GB of files stolen from more than 200 US police departments and fusion training centers.



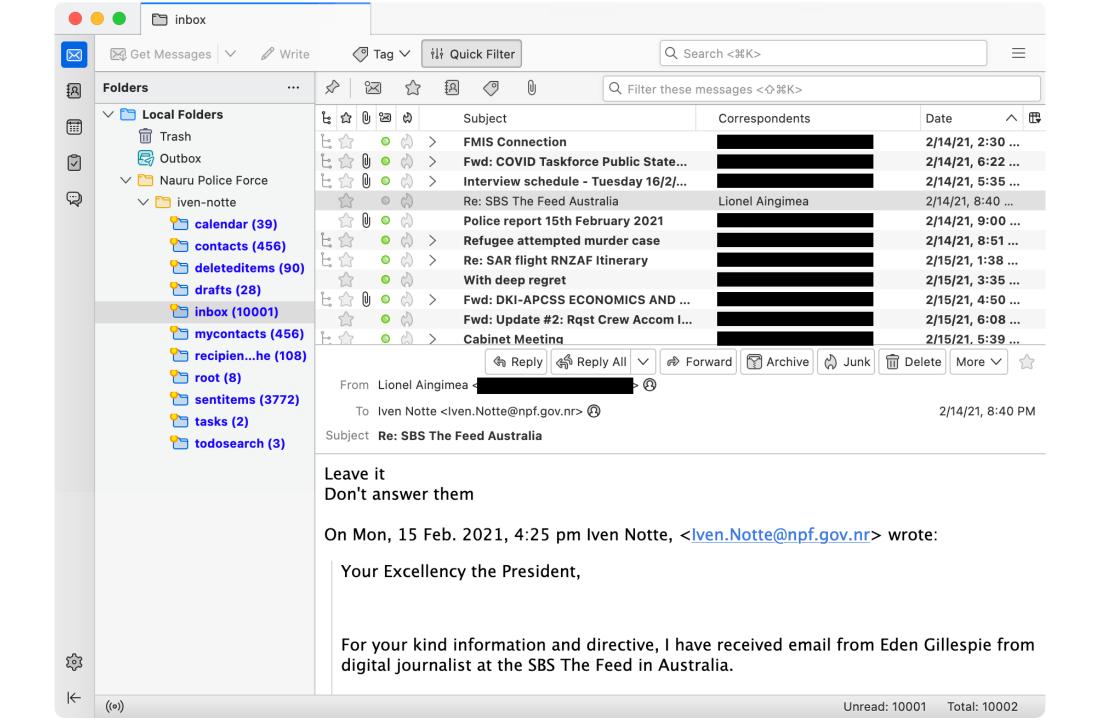
Written by Catalin Cimpanu, Contributor on July 7, 2020



German authorities have seized today a web server that hosted BlueLeaks, a website that provided access to internal documents stolen from US police departments.

The server belonged to DDoSecrets (Distributed Denial of Secrets), an activist group that published the files last month, in mid-June.

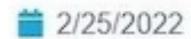
### Reading Other People's Email



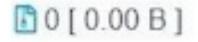
### Data Analysis with Python

#### "WARNING"

The Conti Team is officially announcing a full support of Russian government. If any body will decide to organize a cyberattack or any war activities against Russia, we a re going to use our all possible resources to strike back at the critical infrastructures of an enemy.







```
"_id": "FmFZbde9ACs3gtw27",
 "rid": "GENERAL",
 "msg": "Некоторые американские сенаторы предлагают помимо соцсетей блокировать в России ещё и PornHub!",
 "ts": "2022-02-24T22:02:38.276Z",
 "u" - 1
  " id": "NKrXj9edAPWNrYv5r",
  "username": "thomas",
  "name": "thomas"
 },
 "urls": [],
 "mentions": [],
 "channels":
 "md"
    "type": "PARAGRAPH",
    "value":
       "type": "PLAIN TEXT",
       "value": "Некоторые американские сенаторы предлагают помимо соцсетей блокировать в России ещё и PornHub!"
 " updatedAt": "2022-02-24T22:02:38.293Z"
```

```
>>> with open("2022-02-24-general.json") as f:
       data = json.load(f)
>>> for message in data["messages"]:
       print(f"{message['ts']} {message['u']['username']}: {message['msg']}")
--snip--
2022-02-24T22:02:49.448Z thomas: последние радости у нас заберут
2022-02-24T22:02:44.463Z thomas: ну все, приплыли)
2022-02-24T22:02:38.276Z thomas: Некоторые американские сенаторы предлагают помимо соцсетей
блокировать в России ещё и PornHub!
2022-02-24T22:00:00.347Z thomas:
2022-02-24T21:58:56.152Z rags: угу :(
```

>>> import json

--snip--

### BlueLeaks and the CSV File Format

When I grepped the CSV files in the ncric folder for the word antifa, I found that there were only a handful of references in the files *EmailBuilder.csy*, *Requests.csy*, *SARs.csy*, and *Survey.csy*. In particular, this row in *SARs.csy* stood out because it referenced a student protester, allegedly a member of an antifa group, and mentioned Radicalization/Extremism:

```
micah@trapdoor ncric % grep -ri antifa *.csv
--snip--
SARs.csv:14277, "06/05/20 14:20:09", "6/5/2020", "Marin", "The attached letter was received via US
Postal Service this morning. The letter was passed on from an anonymous party claiming to be a
lawyer who was contacted by [redacted name] who is a University of Oregon student. [Redacted
name appears to be a member of the Antifa group and is assisting in planning protesting efforts
in the Bay Area despite living in Oregon.", "[redacted IP
address]",,"NCRICLawEnforcementReporting",,"Unknown",,"[redacted phone number]","f14e1d15-a052-
489c-968b-5fd9d38544e1","20200596","0820",,"Bay Area",,0,,0,0,0,0,0,0,0,0,0,0,0,0,","[redacted
name]",,,,0,,,,,"[redacted name]","[redacted name]","[redacted name]",,,"Marin County District
Attorney's Office",,,,,"SARF100014\277.pdf",,,,,"- Other
-",,,,,,"Letter.pdf",,,,,,,"[redacted]@marincounty.org","AM","1",,,,,0,0,"Radicalization/
Extremism, Suspicious Incident",, "Emergency Services, Government Facility",,, "No"
--snip--
```

Column Header	Content
SARSid	14277
FormTimeStamp	06/05/20 14:20:09
IncidentDate	6/5/2020
ThreatActivity	Radicalization/Extremism,Suspicious Incident
BriefSummary	The attached letter was received via US Postal Service this morning. The letter was passed on from an anonymous party claiming to be a lawyer who was contacted by [redacted name] who is a University of Oregon student. [Redacted name] appears to be a member of the Antifa group and is assisting in planning protesting efforts in the Bay Area despite living in Oregon.
Subjects	[redacted name]
AgencyOrganizatio nNameOther	Marin County District Attorney's Office
File1	SARF100014\277.pdf
File1Name	Letter.pdf
EmailAddress	[redacted]@marincounty.org

PLEASE SEE THE ATTACHED SOLICITATION I RECEIVED FROM AN ANTIFA
TERRORIST WANTING MY HELP TO BAIL HER AND HER FRIENDS OUT OF JAIL, IF
ARRESTED FOR RIOTING.

MY APOLOGIES FOR NOT PROVIDING MY RETURN E-MAIL AND IDENTITY. I

AM AN ATTORNEY AND CANNOT RISK THIS PIECE OF SHIT ANTIFA – SEE

ATTACHED – FILING A BAR COMPLAINT AGAINST ME. AS YOU CAN CLEARLY SEE

IN THE E-MAIL SHE SENT ME, SHE WAS CLEARLY CROSSING STATE LINES FOR

THE PURPOSES OF RIOTING.

SHE TOLD ME – WHEN I SPOKE WITH HER & AT NO TIME DID I TELL IT WAS A CONFIDENTIAL CONVERSATION – THAT SHE HAD CONTACTED HUNDREDS OF BAY AREA LAWYERS FOR ASSEMBLE THIS LIST.

I NEED YOU TO UNDERSTAND THAT THE SAN FRANCISCO PUBLIC DEFENDERS WILL VIGOROUSLY DEFEND THESE TERRORISTS.

SO YOU WANT TO COORDINATE WITH THE SFPD AND D.A. OFFICE TO SIT IN ON COURT DOCKETS WITH LOOTING AND OTHER RIOTING RELATED CASES BECAUSE YOU WILL THEN FIND MANY OF THESE OUT OF STATE ASSHOLES. HAPPY HUNTING.

I AM DOING THIS BECAUSE I LOVE THE U.S.A. AND I MEANT IT.

Name:

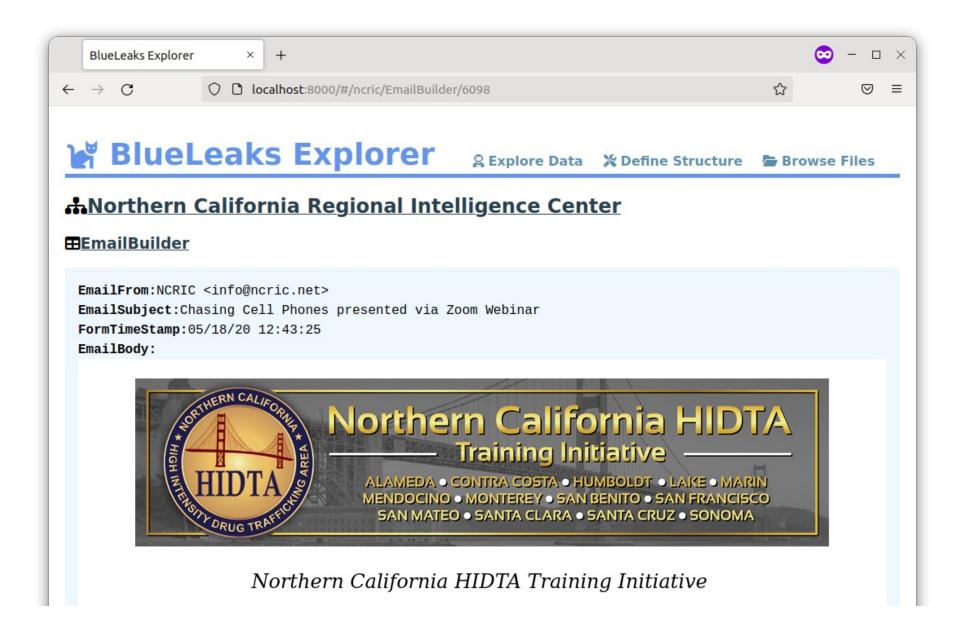
Email Address:

Phone:

Brief description of your legal issue: Hello! I hope this email finds you in good regards.

I am currently working on a list of resources for myself, my friends, and other individuals who are protesting, to refer to if they were to be arrested while protesting. I am mostly catering toward areas in which most of the students who attend my college are from. I am worried that Trump's latest remarks, regarding the new designation naming Antifa as terrorists, will allow for law enforcement to arrest peaceful protesters under the guise of them, "terrorizing," as Antifa is not an actual organization with collective members and could easily be anyone.

is there anyway that I could add your firm, or consenting lawyers under your firm, to a list of resources who will represent protesters pro bono if they were/are to be arrested? Thank you very much for your time.



https://github.com/micahflee/blueleaks-explorer

# JSON, January 6, and Mapping GPS Coordinates





**TECH BY VICE** 

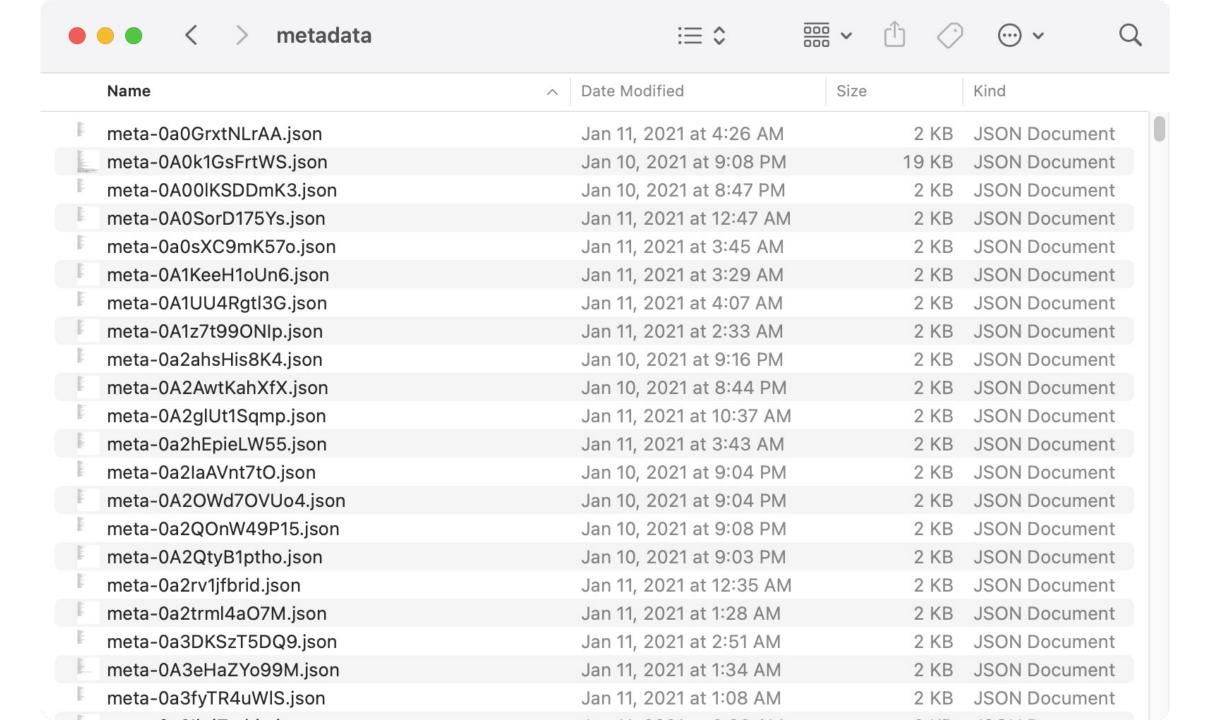
### The Hacker Who Archived **Parler Explains How She Did It (and What Comes** Next)

The hacker, donk\_enby, explained that she only scraped what was publicly available: "I hope that it can be used to hold people accountable and to prevent more death."









You can pipe JSON data into the jq command to indent it and show syntax highlighting in your terminal. For example, try running this command:

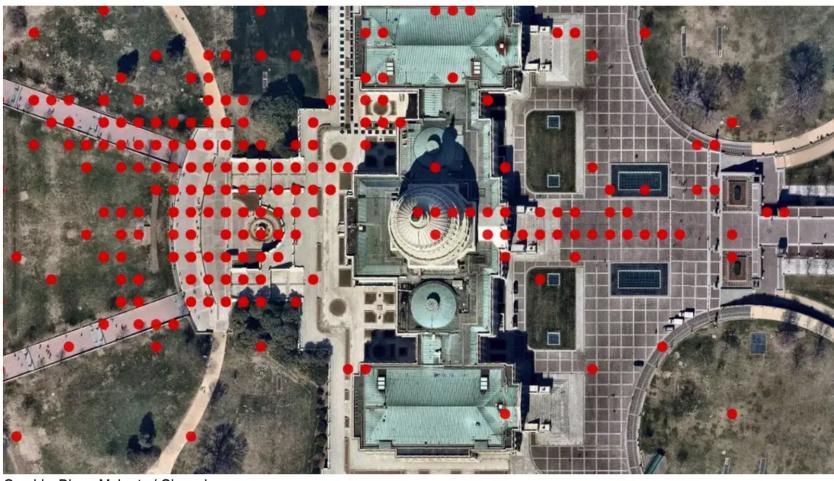
```
cat metadata/meta-HS34fpbzqg2b.json | jq
```

The first part of the command, cat metadata/meta-HS34fpbzqg2b.json, outputs the content of that JSON file, which contains the metadata for a single Parler video. The second part, | jq, pipes that output as input into jq. The output of this command in your terminal should look like this:

```
"SourceFile": "-",
   "ExifToolVersion": 12,
    "FileType": "MOV",
    "FileTypeExtension": "mov",
    "MIMEType": "video/quicktime",
    "MajorBrand": "Apple QuickTime (.MOV/QT)",
    "MinorVersion": "0.0.0",
    "CompatibleBrands": [
      "qt
--snip--
    "GPSLatitude": "38 deg 53' 26.52\" N",
    "GPSLongitude": "77 deg 0' 28.44\" W",
    "Rotation": 180,
   "GPSPosition": "38 deg 53' 26.52\" N, 77 deg 0' 28.44\" W"
```

## Parler Users Breached Deep Inside U.S. Capitol Building, GPS Data Shows

By Dell Cameron and Dhruv Mehrotra | Published January 12, 2021 | Comments (193)

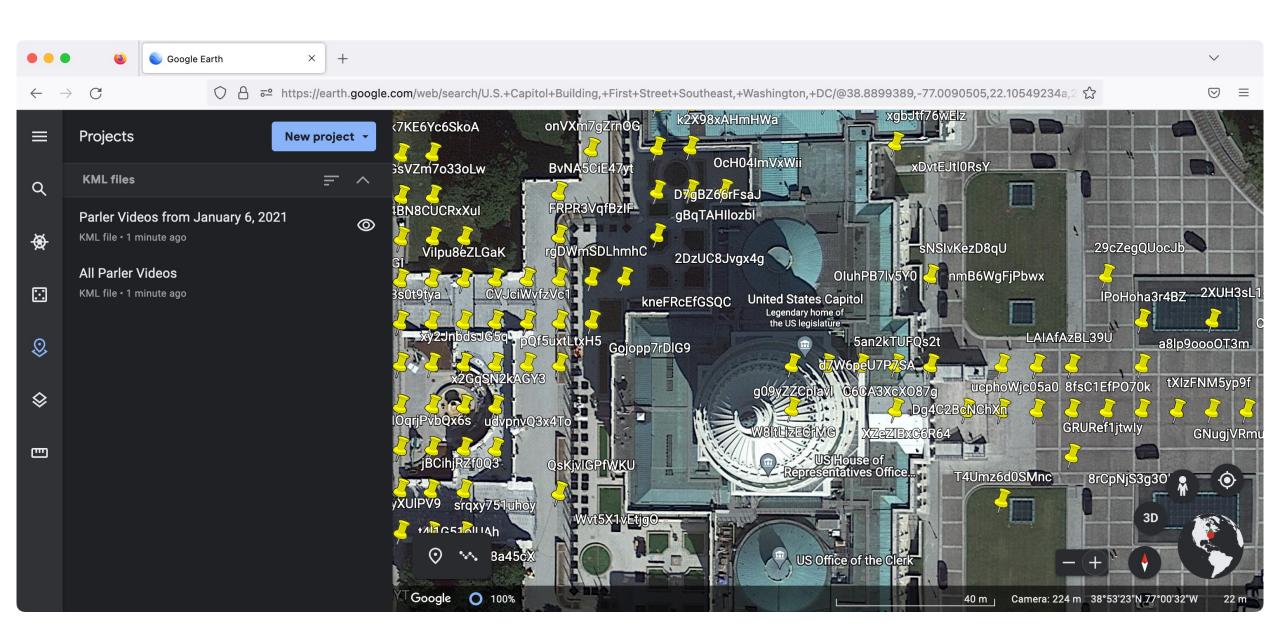


Graphic: Dhruv Mehrota / Gizmodo

```
\square ...
EXPLORER
                                       homework-11-4.py 1 ×
      ∨ HACKS-LEAKS-AND-REVELATIONS
                                       chapter-11 > • homework-11-4.py > ...
                                              import click
        > .vscode
                                              import os
        > appendix-a
                                              import json
 وړ
        > appendix-b
                                              import simplekml
        > chapter-1
        > chapter-2
                                              def json_filename_to_parler_id(json_filename):
        > chapter-3
                                                   return json_filename.split("-")[1].split(".")[0]
        > chapter-4
 > chapter-5
                                         10
        > chapter-6
                                         11
                                              def gps_degrees_to_decimal(gps_coordinate):
                                         12
                                                   parts = gps_coordinate.split()
        > chapter-7
                                         13
                                                  degrees = float(parts[0])
        > chapter-8
                                         14
                                                  minutes = float(parts[2].replace("'", ""))
        > chapter-9
                                         15
                                                  seconds = float(parts[3].replace('"', ""))
        > chapter-10
                                                  hemisphere = parts[4]

∨ chapter-11

                                         17
                                                  gps_decimal = degrees + (minutes / 60) + (seconds / 3600)
                                                   if hemisphere == "W" or hemisphere == "S":
         .gitignore
                                         19
                                                       qps decimal *= -1
         homework-11-1-filenames.p...
                                         20
                                                   return gps_decimal
         homework-11-1.md
                                         21
         homework-11-2.md
                                         22
                                         23
                                              @click.command()
         homework-11-2.py
                                         24
                                              @click.argument("parler_metadata_path")
         homework-11-3.md
                                         25
                                              def main(parler_metadata_path):
         homework-11-3.py
                                                   """Create KML files of GPS coordinates from Parler metadata"""
         homework-11-4.md
                                         27
                                                  kml all = simplekml.Kml()
         homework-11-4.py
                                                  kml_january6 = simplekml.Kml()
                                         29
        > chapter-12
                                         30
                                                   for filename in os.listdir(parler_metadata_path):
        > chapter-13
                                         31
                                                       abs_filename = os.path.join(parler_metadata_path, filename)
        .gitignore
                                         32
                                                       if os.path.isfile(abs filename) and abs filename.endswith(".ison"):
```

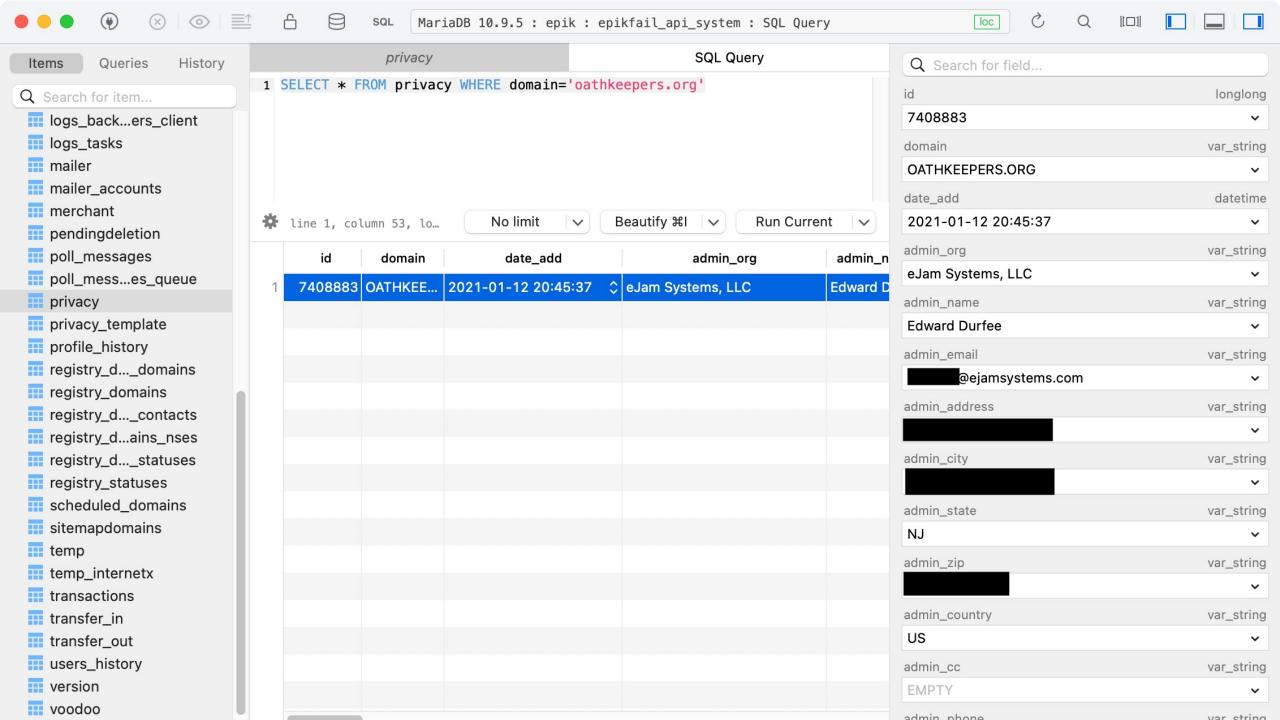


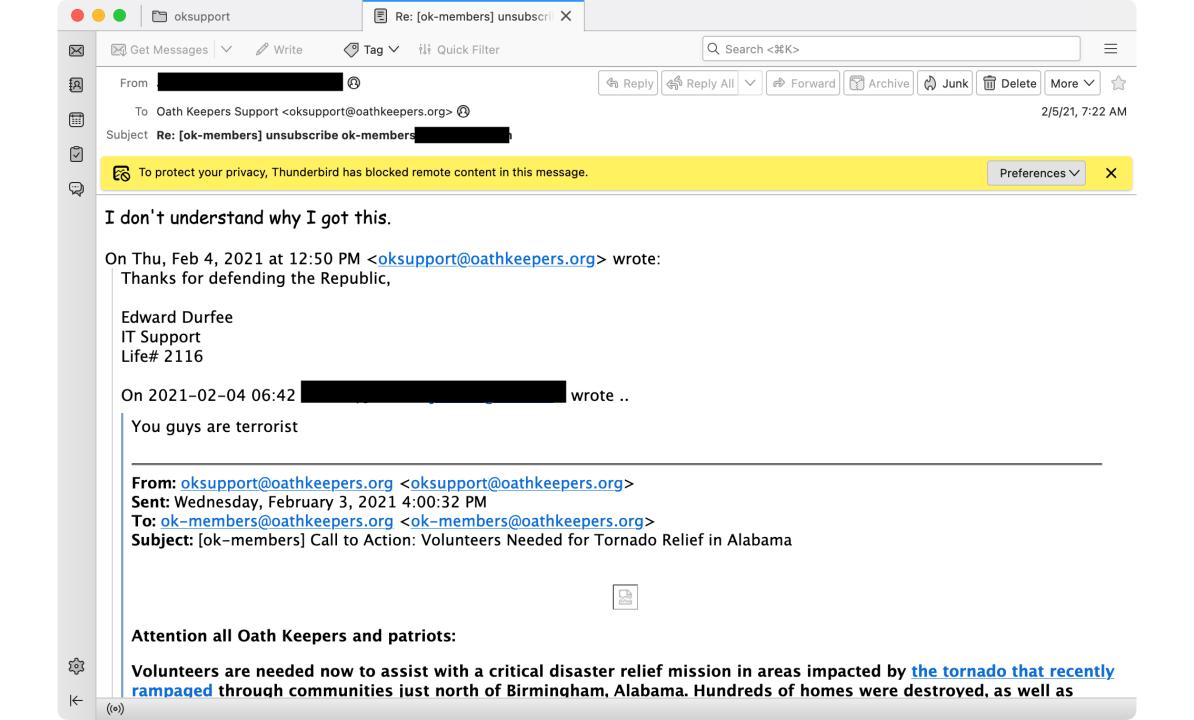
# SQL, Epik Fail, and Extremism Research

### Whois Record for OathKeepers.org

#### - Domain Profile

Registrant	REDACTED FOR PRIVACY	
Registrant Org	Anonymize, Inc.	
Registrant Country	US	
Registrar	Epik Inc. IANA ID: 617 URL: https://www.epik.com Whois Server: http://whois.epik.com abuse@epik.com (p) +1.425366881	
Registrar Status	ok	
Dates	5,137 days old Created on 2009-03-01 Expires on 2033-03-01 Updated on 2023-03-08	*
Name Servers	NS1.DREAMHOST.COM (has 1,356,935 domains) NS2.DREAMHOST.COM (has 1,356,935 domains)	<b>~</b>





## Pandemic Profiteers and COVID-19 Disinformation

#### The Intercept\_

# NETWORK OF RIGHT-WING HEALTH CARE PROVIDERS IS MAKING MILLIONS OFF HYDROXYCHLOROQUINE AND IVERMECTIN, HACKED DATA REVEALS

The data also reveals that 72,000 people paid at least \$6.7 million for Covid-19 consultations promoted by America's Frontline Doctors and vaccine conspiracist Simone Gold.

Micah Lee

September 29 2021, 6:37 a.m.





### HOUSE CORONAVIRUS COMMITTEE LAUNCHES INVESTIGATION INTO ORGANIZATIONS PUSHING HYDROXYCHLOROQUINE, IVERMECTIN

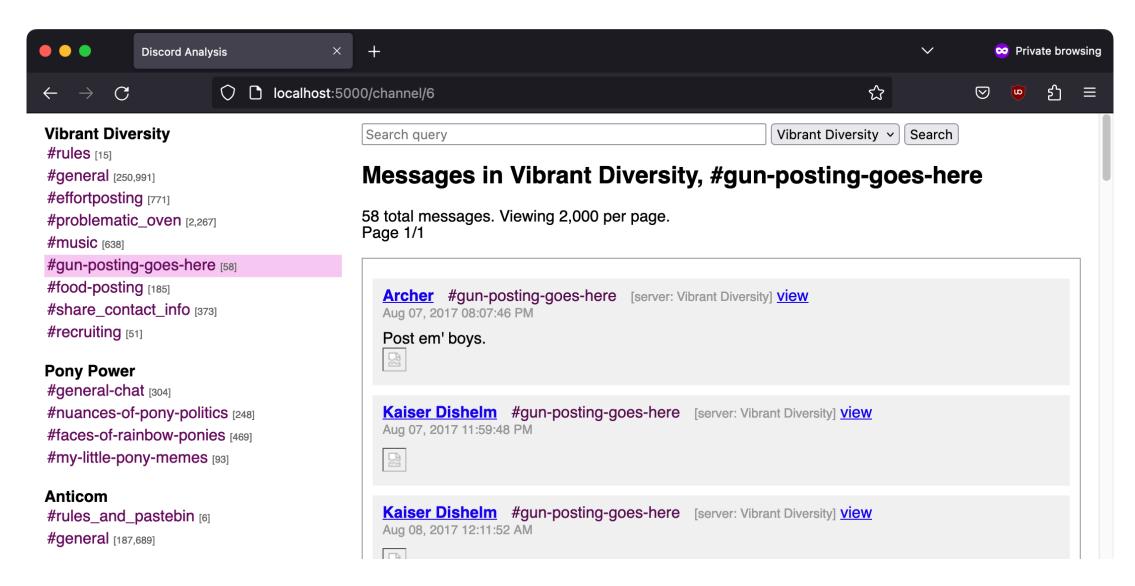
The investigation into America's Frontline Doctors and SpeakWithAnMD.com comes after an Intercept story revealed a right-wing network making millions.

Micah Lee

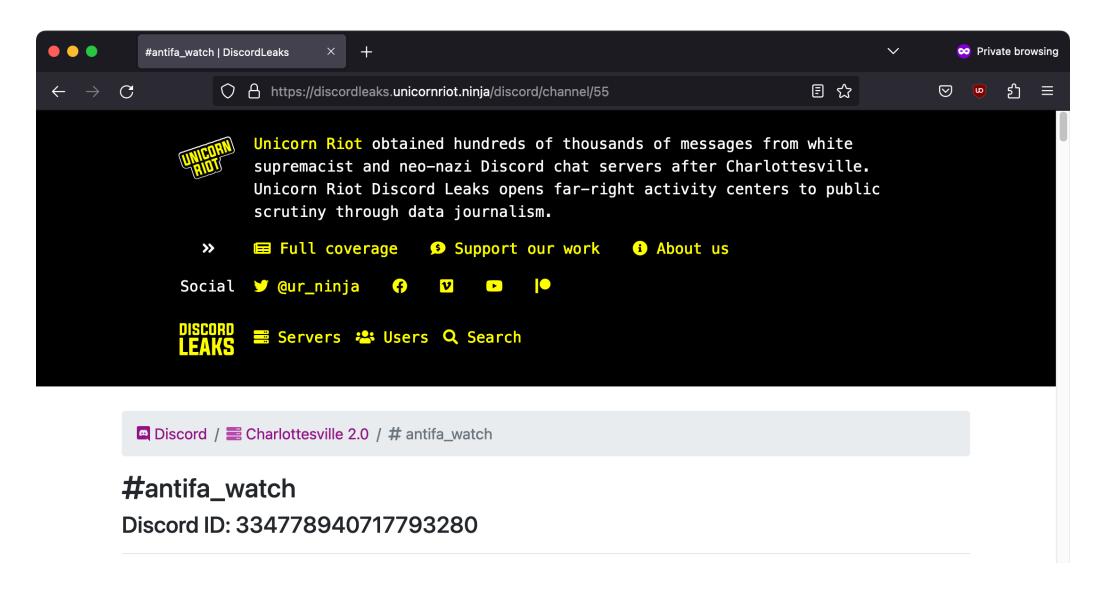
November 2 2021, 4:46 a.m.

### Neo-Nazis and Their Chat Rooms





https://github.com/micahflee/discord-analysis



https://discordleaks.unicornriot.ninja/

# Jury Finds Rally Organizers Responsible for Charlottesville Violence

Jurors found the main organizers of the deadly far-right rally in Charlottesville, Va., in 2017 liable under state law, awarding more than \$25 million in damages, but deadlocked on federal conspiracy charges.

ありがとうございます! I'm Micah Lee, micah@micahflee.com

Hacks, Leaks, and Revelations to be released ~September 2023

pre-order from No Starch Press:

https://nostarch.com/hacks-leaks-and-revelations