IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF COLUMBIA

<table>
<tr><td>

NATIONAL VETERANS LEGAL
SERVICES PROGRAM,
NATIONAL CONSUMER LAW
CENTER, and ALLIANCE FOR
JUSTICE, for themselves and all
others similarly situated,
<div align="center">*Plaintiffs,*</div>

v.

UNITED STATES OF AMERICA,
<div align="center">*Defendant.*</div>

</td><td>

Case No. 16-745

</td></tr>
</table>

DECLARATION OF THOMAS LEE AND MICHAEL LISSNER

Thomas Lee and Michael Lissner hereby declare as follows:

Thomas Lee Background and Experience

1.   Thomas Lee is a software developer and technologist with a background in federal government transparency issues. He currently develops software for a large venture-backed software company. In this capacity he uses cloud-based storage and computation services on a daily basis and assists in cost estimation, planning and optimization tasks concerning those services.

2.   Before taking on his current private-sector role in 2014, Mr. Lee spent six years working at the Sunlight Foundation, serving four of those years as the Director of Sunlight Labs, the Foundation's technical arm. The Sunlight Foundation is a research and advocacy

<div align="center">1</div>

organization focused on improving government transparency. Sunlight Labs' work focused on the modernization of government information technology and improving the distribution of government data. This work included technical project management, budgeting, media appearances and testimony before Congress, among other tasks.

3.   Prior to joining the Sunlight Foundation, Mr. Lee built websites for large nonprofits, the U.S. Navy, and the offices of individual members and committees within the U.S. Senate and House of Representatives. Mr. Lee's resume is attached to this declaration.

Michael Lissner Background and Experience

4.   Michael Lissner is the executive director of Free Law Project, a nonprofit organization established in 2013 to provide free, public, and permanent access to primary legal materials on the internet for educational, charitable, and scientific purposes to the benefit of the general public and the public interest. In this capacity he provides organizational management, publishes advocacy materials, responds to media inquiries, and writes software.

5.   Since 2009, Free Law Project has hosted RECAP, a free service that makes PACER resources more widely available. After installing a web browser extension, RECAP users automatically

contribute PACER documents they purchase to a central repository. In return, when using PACER, RECAP users are notified if a document exists in the RECAP central repository. When it does, they may download it directly from the RECAP repository, avoiding the need to pay PACER fees.

6.   In the course of maintaining and improving RECAP, Mr. Lissner has become extensively familiar with PACER. During this time RECAP's archive of PACER documents has grown to more than 1.8 million dockets containing more than 40 million pages of PACER documents.

7.   Mr. Lissner has conducted extensive research on the operation and history of the PACER system. Among other topics, this research has focused on the costs of PACER content and the history of PACER fees. This research is available on the Free Law Project website.[1] Mr. Lissner's resume is attached to this declaration.

Expert Assignment and Materials Reviewed

8.   We have been asked by the plaintiffs' counsel in this case to evaluate the reported fee revenue and costs of the PACER system in light of our knowledge of existing information technology and data-storage costs, our specific knowledge of the PACER system, and our background in federal government information systems.

---

[1] https://free.law/pacer-declaration/

9.   Specifically, the plaintiffs' counsel have asked us to offer an opinion on whether the Administrative Office of the U.S. Courts (AO) is charging users more than the marginal cost of disseminating records through the PACER system—in other words, to use the language of the E-Government Act of 2002, the "expenses incurred in providing" access to such records for which it is "necessary" to charge a fee "for [the] services rendered."

10.   In forming our opinion, we have reviewed the Plaintiffs' Statement of Undisputed Material Facts and some of the materials cited in that statement, including a spreadsheet provided to the plaintiffs' counsel in discovery (Taylor Decl., Ex. L) and the Defendant's Response to Plaintiffs' First Set of Interrogatories (Taylor Decl., Ex. M).

11.   We also rely upon our accumulated experience as technologists and government transparency advocates.

Reasoning and Conclusions on Marginal Cost

12.   As we explain in detail below, it is overwhelmingly likely that the PACER system, as operated by the Administrative Office of the Courts (AO), collects fees far in excess of the costs associated with providing the public access to the records it contains.

13.   The following calculations are intended to convey fair but approximate estimates rather than precise costs.

14.   The marginal cost of providing access to an electronic record consists of (a) the expenses associated with detecting and responding to a request for the record; (b) the bandwidth fees associated with the inbound and outbound transmissions of the request and its response; and (c) the pro rata expense associated with storing the records in a durable form between requests.

15.   As a point of comparison we use the published pricing of Amazon Web Services (AWS). AWS leads the market for cloud computing services[2] and counts organizations including Netflix, Adobe Systems, and NASA among its customers. Like most cloud providers, AWS pricing accounts for complex considerations such as equipment replacement, technical labor, and facilities costs. Although the division is profitable, AWS prices are considered highly competitive. AWS services are organized into regions, each of which represents a set of data centers in close geographic and network proximity to one another.

16.   For our evaluation, we first consider the cost of storage. Researcher Matthew Komorowski[3] and data storage firm BackBlaze[4] have published storage cost time series that when combined cover the period dating from the PACER system's 1998 debut to the present.

---

[2]
https://www.srgresearch.com/articles/leading-cloud-providers-continue-run-away-market.
[3] http://www.mkomo.com/cost-per-gigabyte
[4] https://www.backblaze.com/blog/hard-drive-cost-per-gigabyte/

During this time their data shows the cost of a gigabyte of storage falling from $65.37 to $0.028, a reduction of over 99.9%. During this same time period PACER's per-page fees increased 43%, from $0.07 to $0.10.

17.   The effect of economies of scale makes it difficult to assemble comparable time series for bandwidth and computing costs. We are therefore unable to easily compare PACER fees' growth rate to the change in bandwidth and computing costs from 1998 to the present.

18.   Fortunately, it is possible to compare recent PACER fee revenue totals to reasonable contemporary costs for the technical functionality necessary to perform PACER's record retrieval function. The AWS Simple Storage Service (S3) provides this necessary data storage and retrieval functionality and publishes straightforward and transparent pricing for it. S3 costs vary by region. Using the prices published on August 27, 2017 for the "GovCloud" region, which is designed for U.S. government users, we find storage prices of $0.039 per gigabyte[5] per month for the first 50 terabytes, $0.037 per gigabyte per month for the next 450 terabytes, and $0.0296 per gigabyte per month for the next 500 terabytes. Retrieving an item from the

---

[5] The quantity of data contained in a terabyte/gigabyte/megabyte/kilobyte varies slightly according to which of two competing definitions is used. Our analysis employs the definitions used by Amazon Web Services. c.f. https://docs.aws.amazon.com/general/latest/gr/glos-chap.html

GovCloud region currently costs $0.004 per 10,000 requests, plus data transmission at $0.01 per gigabyte.

19.   Determining how these prices might apply to PACER's needs requires knowledge of the PACER system's size. We are not aware of a current and authoritative source for this information. Instead, we employ an estimate based on two sources from 2014: that year's Year-End Report on the Federal Judiciary,[6] and an article published in the *International Journal for Court Administration.*[7] The former states that PACER "currently contains, in aggregate, more than one billion retrievable documents." The latter states that the PACER "databases contain over 47,000,000 cases and well over 600,000,000 legal documents; approximately 2,000,000 new cases and tens of millions of new documents are entered each year." Although the large difference in document counts makes it unlikely that both of these estimates are correct, they provide an order of magnitude with which to work. For the sake of our estimate we double the larger of these numbers and make the generous assumption that PACER now contains two billion documents.

20.   Mr. Lissner's custodianship of the RECAP archive allows us to make estimates of the typical properties of PACER documents.

---

[6] https://www.supremecourt.gov/publicinfo/year-end/2014year-endreport.pdf
[7] Brinkema, J., & Greenwood, J.M. (2015). E-Filing Case Management Services in the US Federal Courts: The Next Generation: A Case Study. International Journal for Court Administration, 7(1). Vol. 7, No. 1, 2015.

21.   The RECAP Archive contains the most-requested documents from PACER, making them appropriate for our analysis.

22.   Mr. Lissner finds an average document size of 254 kilobytes and 9.1 pages, and therefore an average page size of 27.9 kilobytes. Assuming a PACER database size of two billion documents and the prices recorded above, we calculate that annual storage costs of the the PACER database on S3 would incur fees totaling $226,041.60.

23.   This leaves the task of estimating the costs incurred by the retrieval of documents. To do this we must estimate the total number of requests served by PACER each year. The PACER fee revenue reported for 2016 in the spreadsheet provided to the plaintiffs' counsel in discovery is $146,421,679. The per-page PACER fee in 2016 was $0.10. Simple arithmetic suggests that approximately 1,464,216,790 pages were retrieved from PACER in 2016.

24.   This calculation does not reflect the 30 page/$3.00 per-document cap on fees built into PACER's price structure; nor the fact that some of the revenue comes from search results, which are also sold by the page; nor any other undisclosed discounts.

25.   The RECAP dataset's 9.1 page average document length suggests that the fee cap might not represent a substantial discount to users in practice.

27.   Out of an abundance of caution against underestimating costs, we account for these inaccuracies by rounding the estimated request count up to two billion for the following calculations.

28.   Using aforementioned S3 prices for retrieving an item from storage, this volume of annual requests would incur $800 in fees. An additional $558.24 in bandwidth costs would also be incurred. This yields a total yearly estimate for storing and serving PACER's dataset using AWS S3's GovCloud region of $227,399.84, or 0.16% of PACER's reported 2016 fee revenue.

29.   The tremendous disparity between what the judiciary actually charges in PACER fees and what is reasonably necessary to charge is illustrated by two alternative calculations. The first considers what the per page fee could be if PACER was priced according to our calculations. Including storage costs, we estimate that the per page cost of retrieving a document from PACER could cost $0.0000006 (about one half of one ten-thousandth of a penny). The second alternate calculation considers how many requests PACER could serve if the fees it currently collects were used exclusively and entirely for providing access to its records. Assuming no change in the size of the dataset and using the storage costs calculated in association with that size, $146,195,637.40 in fee revenue remains to cover document requests and bandwidth. At the previously cited rates, this would

cover the costs associated with serving 215,271,893,258,900 requests, or approximately 1,825 pages per day for every person in the United States.

Reasoning and Conclusions on Reasonableness of Costs

30.   We offer the preceding analysis with three caveats. First, at the time of PACER's design and implementation, cloud computing services were not widely available and the cost savings associated with their scale could not be achieved. It is therefore reasonable to assume that PACER's costs could be artificially high due to the time in which it was built, although effective ongoing maintenance and modernization should attenuate this effect. Second, although the Administrative Office of the Courts could directly use the Amazon Web Services we discuss, it would not be uncommon or unreasonable to purchase those services through a reseller who increases their price by some amount. Third, it is important to note that as outside analysts with limited information, we cannot anticipate or account for all of the costs that could conceivably be associated with access to PACER records.

31.   But it is noteworthy that PACER fees increased during a period of rapidly declining costs in the information technology sector. Even after taking the preceding caveats into account, we are unable to offer a reasonable explanation for how PACER's marginal cost for

serving a record could be many orders of magnitude greater than the contemporary cost of performing this function.
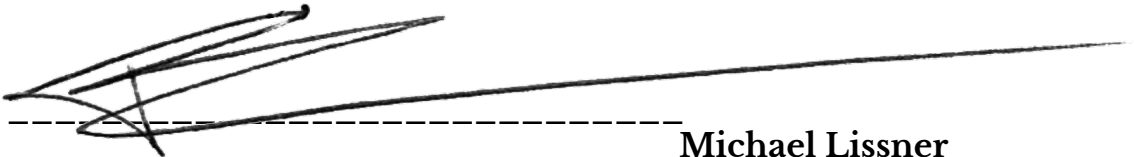
32.   It is overwhelmingly likely that the PACER system, as administered by the AO, collects fees far in excess of the costs associated with providing the public access to the records it contains.

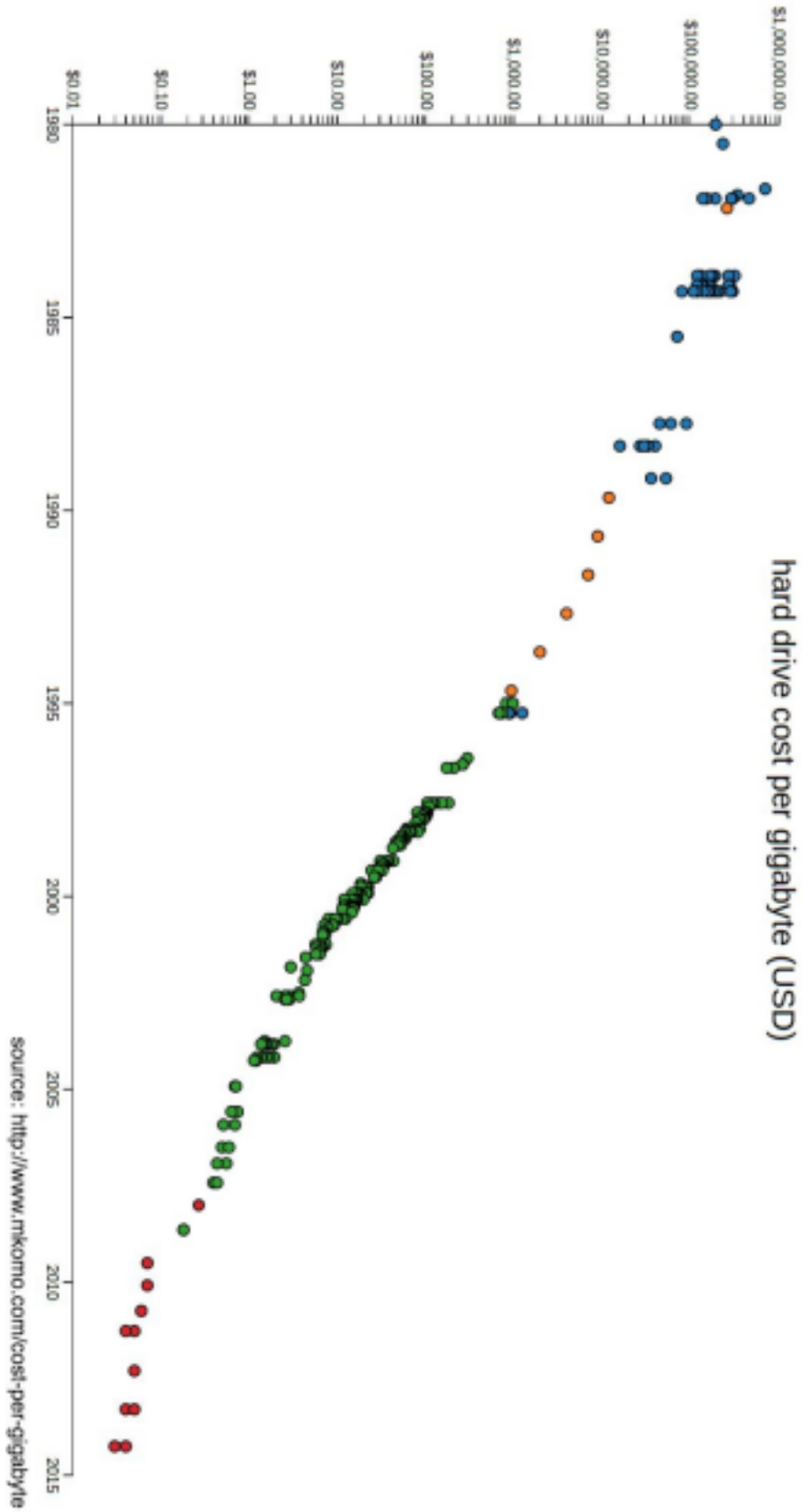33.   We declare under penalty of perjury that the foregoing is true and correct.

Executed on August 28, 2017.

_____
                                              Thomas Lee

_____
                                              Michael Lissner

hard drive cost per gigabyte (USD)

source: http://www.mkomo.com/cost-per-gigabyte

Backblaze Average Cost per GB for Hard Drives

By Quarter: Q1 2009 - Q2 2017
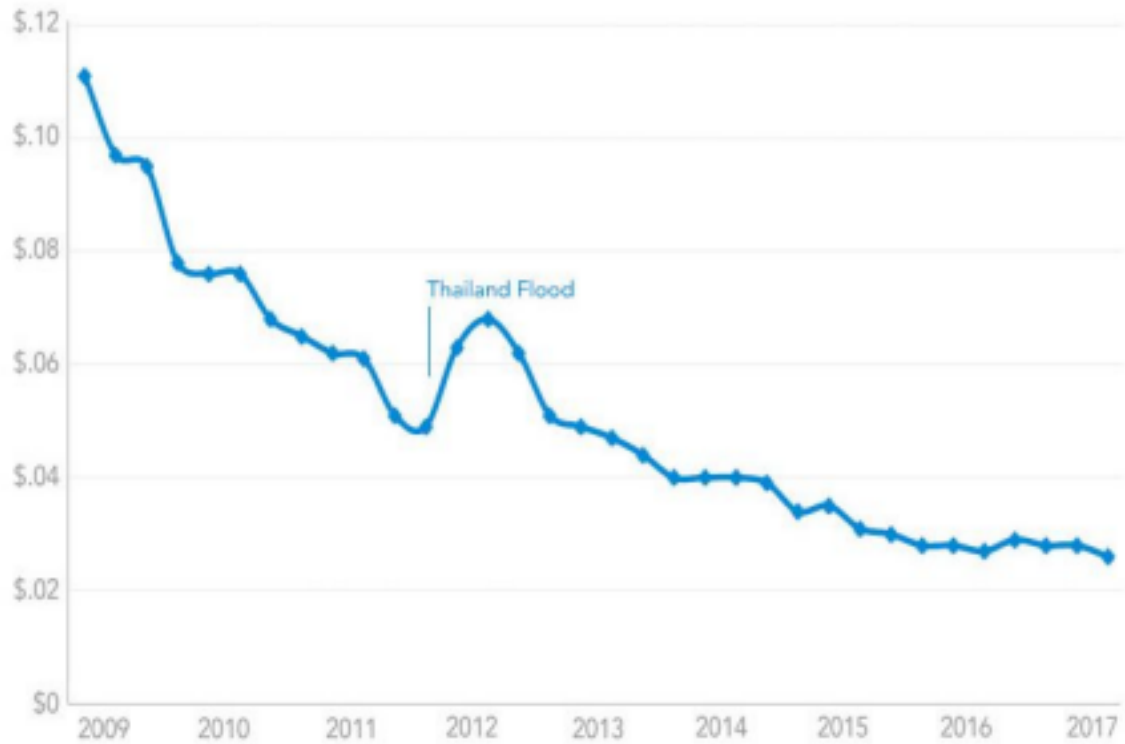
source: https://www.backblaze.com/blog/hard-drive-cost-per-gigabyte/

Thomas Lee

understanding / making / explaining technology
https://www.linkedin.com/in/tom-lee-a2112387/

50 Q St NE #2
Washington, DC 20002
(703) 944-7654
thomas.j.lee@gmail.com
https://github.com/sbma44

EXPERIENCE

Mapbox — *Geocoding Lead*

JUNE 2010 - PRESENT

Guided Mapbox's location search team through a period of fast growth and into commercial success. Also performed a variety of legal, security and hardware tasks.

– Oversaw growth of geocoding business from 1% to 21% of revenue by line item, 39% to 71% by related-deal revenue. Shipped code, performed sales engineering, led hiring, participated in enterprise support, evaluated & managed compliance for licensed data.

– Managed federal government relations, including Congressional lobbying & testimony, agency meetings & writing op-eds on behalf of leadership. Liaised with relevant open data communities.

– Coordinated outside counsel during patent defense.

– Designed and implemented royalty tracking pipeline and mobile SDK battery test methodology. Assisted in design of mobile telemetry security systems. Authored first version of security protocols for participation in infosec events with hostile networks.

Sunlight Foundation — *CTO*

DECEMBER 2008 - JUNE 2010

Managed Sunlight Labs' twenty-two person technology department during its prime years of influence and size.

– Conceived, planned and executed mission-oriented technology projects.

– Represented Sunlight's positions on various government transparency measures in Congressional testimony, speaking engagements, writing, and media appearances.

– Expanded historically web dev-focused team to include political scientists, journalists, data analysts & mobile app developers.

– Primary author of grants and reports for bulk of Sunlight funding.

– Evaluated grant applications for potential funding. Managed relationships with peer organizations, funders and grantees.

SKILLS

writing · team management · software development · data analysis · speaking · system administration · information security · embedded systems

TECHNOLOGIES

Expert
Javascript / Node.js · Python / Django / Flask · SQL / PostgreSQL · bash / GNU · Docker · AWS / EC2 / ECS / CloudFormation / DynamoDB / ElastiCache / Kinesis / S3 · PHP / Drupal / Wordpress · AVR / Arduino · QGIS · GDAL · PostGIS · Mapbox

Productive
Perl · Ruby · HTML5 · CSS

Tourist
C · C++ · Swift/XCode · three.js

ORGANIZATIONS

OpenAddresses · FLOC · HacDC · DCist

EchoDitto — *Sr. Software Architect*

DECEMBER 2005 - DECEMBER 2008

Designed & implemented LAMP applications for campaigns and large nonprofits, primarily using the Drupal and WordPress frameworks.

- Assisted in requirement-gathering, copy editing and writing, strategy brainstorming, customer interaction and visual design.

- Developed variety of reporting mechanisms (SQL/Perl/Ruby).

- Launched, maintained and generated bulk of content for developer-focused EchoDitto Labs site.

Competitive Innovations — *Software Developer*

August 2002 - DECEMBER 2005

Created ASP.NET/Microsoft CMS-backed websites for committees and member offices in the U.S. House of Representatives; the U.S. Navy; George Washington University Law School; Miami Dade Community College; and the Corporate Executive Board.

- Interviewed, evaluated, trained and participated in the management of junior technical staff.

- Possessed security clearance as of December 2005.

SELECTED CLIPS

What Everyone Is Getting Wrong About Healthcare.gov
Wonkblog, Washington Post
http://www.washingtonpost.com/blogs/wonkblog/wp/2013/10/07/what-everyone-is-getting-wrong-about-healthcare-gov/

The Cost of Hashtag Revolution
The American Prospect
http://prospect.org/article/cost-hashtag-revolution

The Deleted Tweets of Politicians Find a New Home
Tell Me More (NPR)
http://www.npr.org/2012/06/06/154432624/the-deleted-tweets-of-politicians-find-a-new-home

Enhancing Accountability and Increasing Financial Transparency
U.S. Senate Committee on the Budget
https://www.budget.senate.gov/hearings/enhancing-accountability-and-increasing-financial-transparency

EDUCATION

University of Virginia — *BA, Cognitive Science*

1998-2002

Concentration in neuroscience, with work in the Levy Computational Neuroscience Lab. Computer Science minor. Echols Scholar.

MICHAEL JAY LISSNER

mike@free.law • (909) 576-4123 • 2121 Russell St., Suite B, Berkeley, CA 94705

EXPERIENCE

Executive Director and Lead Developer   **2013-Present**
Free Law Project   **Emeryville, CA**
Founded Free Law Project as a 501(c)(3) non-profit. My responsibilities as founder/director include identifying and pursuing grants and contracts, handling the marketing and accounting needs of the organization, and developing solutions for our stakeholders.

Free Law Project has been awarded grants or contracts from Columbia University, Georgia State University, University of Baltimore School of Law, and The John S. and James L. Knight Foundation, and has partnered with Google, Inc. and the Center for Internet and Technology Policy at Princeton University.

I am the lead developer for several of Free Law Project's biggest initiatives, including:
- The first ever full-text search interface for documents from the PACER system, containing nearly 20M records;
- The creation of the largest archive of American oral argument recordings, consisting of nearly one million minutes of recordings;
- The development of a comprehensive database of American judges;
- The curation of 4M court opinions, which are available via a powerful search interface, as bulk data, or via the first ever API for legal opinions;
- The creation of a web scraping infrastructure that has gathered more than 1M documents from court websites.

This work has enabled a number of research papers, made legal research more competitive, provided a useful resource to journalists, and helped innumerable people to engage in the legal system.

New Product Designer/Developer   **2012-2013**
Recommind, Inc.   **San Francisco, CA**
- Worked with the new products team to design and develop new enterprise-class products for AMLAW-50 law firms.
- Led design of new API-driven document sharing platform from initial concept to final specification, seeking stakeholder approval from upper management, sales, product management, and development teams. This process was guided by the creation of paper prototypes and low fidelity wireframe diagrams, culminating in high fidelity mock-ups and a written specification.

Solutions Developer   **2010-2012**
Recommind, Inc.   **San Francisco, CA**
- Designed and developed new features, products and processes for internal team of technical consultants.
- Implemented distributed search systems for top international law firms.
- Collaborated with internal and external stakeholders to gather requirements and scope work.
- Developed custom crawlers and search indexes for systems with millions of records.

Technology Intern   **Summer, 2009**
Center for Democracy and Technology   **San Francisco, CA**
Wrote design specification and began implementation of location privacy enhancements for the new Android operating system.

Systems Analyst and Community Researcher   **2005-2008**
Community Services Bureau   **Contra Costa County**
- Designed and implemented system for reporting educational outcomes and program metrics to senior management.
- Researched and wrote federally-mandated annual assessment of community needs.
- Worked with contractors to administer departmental databases and systems.
- Discovered and responsibly-disclosed security vulnerabilities in department systems, protecting tens of thousands of child and parent records.
- Tracked and reported daily enrollment of more than 2,000 children.

E D U C A T I O N   **School of Information, UC Berkeley**   **2008-2010**
- Masters in Information Management and Systems (MIMS), with a focus on Internet Law and Policy and a certificate in Management of Technology from Haas School of Business
- Theoretical coursework in information privacy, policy and economics, intellectual property law, and technology strategy
- Technical coursework in security, networking, programming paradigms, distributed computing, API design, and information architecture
- Taught Web Architecture summer seminar to class of twenty undergraduates including fundamentals of networking, dynamic websites, and browsers

University of California, Berkeley Extension   **2005-2008**
- Unix/Linux fundamentals
- System administration programming, with focus on shell scripting and Python
- Advanced Java programming

Pitzer College, Claremont, California   **2000-2004**
- Bachelor of Arts in English and World Literature with a minor in Spanish Language and Literature
- Coursework in economics, mathematics and C++ programming

P R O J E C T S &   **CourtListener.com**
R E S E A R C H
My capstone project at UC Berkeley and now a core initiative of Free Law Project, CourtListener.com is an open-source legal research tool that provides daily awareness and raw data to users via custom email alerts, Atom feeds, podcasts, a RESTful API, and bulk data. CourtListener currently:
- Hosts the RECAP Archive, a collection of nearly 20M PACER documents;
- Has 4M Boolean-searchable opinions in its corpus;
- Has more nearly 700 days of oral argument audio;
- Has a comprehensive database of American judges;
- Receives thousands of API hits per day;
- Tracks every high court in the country, adding their opinions as they are published.

https://www.courtlistener.com | https://github.com/freelawproject/courtlistener

Seal Rookery

The Free Law Project Seal Rookery is a small project to collect and distribute all government seals in the United States. Currently, the project has more than 200 judicial seals.

https://github.com/freelawproject/seal-rookery

Selected Policy, Legal and Security Papers

- CourtListener.com: A platform for researching and staying abreast of the latest in the law
- Wikipedia.org: Jacobsen v. Katzer, Zeran v. AOL
- The Layered FTC Approach to Online Behavioral Advertising
- Technology Revolution and the Fourth Amendment
- Transparent Panacea: Why Open Email is Fraught with Problems
- Proactive Methods for Secure Design
- Breaking reCAPTCHA
- Facebook's Battle Sign: A Security Analysis

http://michaeljaylissner.com/projects-and-papers/

Additional Websites and Projects

michaeljaylissner.com | free.law | github.com/freelawproject

A D D I T I O N A L

Distance Travel

- **Summer, 2013-2014:** Completed south-bound thru-hike of Te Araroa Trail in New Zealand (2,000 miles). The Te Araroa Trail is considered one of the most-challenging long-distance trails in the world.

- **Summer, 2010:** Completed south-bound bike tour of California coast (1,000 miles).

- **Summer, 2005:** Completed north-bound thru-hike of Pacific Crest Trail from Mexico to Canada via Sierra and Cascade mountains (2,500 miles).