Getting to CSV:
Unlocking Public Data in the Civil Justice Space

Dan Bernstein,
Research Analyst

LSC | America's Partner for Equal Justice
LEGAL SERVICES CORPORATION

Overview

- Why do we care about court data?
- What is civil court data?
- How are we getting court data?
- What lessons can we learn from this project?

What is the Legal Services Corporation (LSC)?

- A congressionally established corporation created to fill “the justice gap”
- The largest funder of civil legal aid for low-income Americans in the United States
- Provided over $400 million in 2018 in funding to 132 grantee organizations in the 50 states and territories

“The Civil Court Data Project”

- Our goal is to gather **state court data** at the **state and local level** to:
  - Answer fundamental questions about the experience of low-income individuals in the civil justice system
  - Develop tools to promote equality in the civil justice system
  - Develop relationships with local organizations who can utilizes court data and analysis

Garnishment in Tennessee during COVID

- Advocates wanted:
    - TN Supreme Court to stop debt collectors from garnishing stimulus checks
    - asked for any recent data on debt collection and garnishment in April 2020
- We determined:
    - nearly 2500 instances of garnishment activity since courts closed
    - $200,000-$400,000 collected in garnishment during COVID
- Our partners submitted the analysis to the Tennessee Supreme Court to inform next steps

What does court data look like?

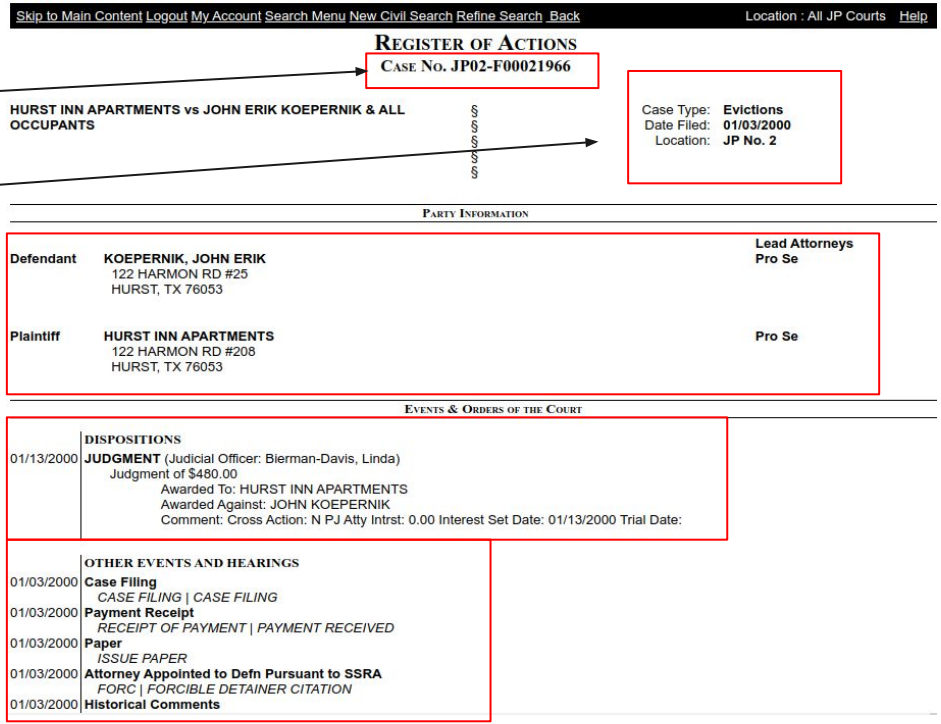
unique case number with the court system

Case:
Case type, date filed, and court location

Parties:
full name, street address, attorney name
('pro se' means no legal representation)

Judgment:
Outcomes, money owed, interest rates, other notes

Events: original filing, payments and fines, hearings



Skip to Main Content Logout My Account Search Menu New Civil Search Refine Search Back Location : All JP Courts Help

REGISTER OF ACTIONS
CASE NO. JP02-F00021966

HURST INN APARTMENTS vs JOHN ERIK KOEPERNIK & ALL OCCUPANTS §§§§§§

Case Type: Evictions
Date Filed: 01/03/2000
Location: JP No. 2

PARTY INFORMATION

Defendant	KOEPERNIK, JOHN ERIK 122 HARMON RD #25 HURST, TX 76053	Lead Attorneys Pro Se
Plaintiff	HURST INN APARTMENTS 122 HARMON RD #208 HURST, TX 76053	Pro Se

EVENTS & ORDERS OF THE COURT

DISPOSITIONS

01/13/2000 **JUDGMENT** (Judicial Officer: Bierman-Davis, Linda)
Judgment of $480.00
Awarded To: HURST INN APARTMENTS
Awarded Against: JOHN KOEPERNIK
Comment: Cross Action: N PJ Atty Intrst: 0.00 Interest Set Date: 01/13/2000 Trial Date:

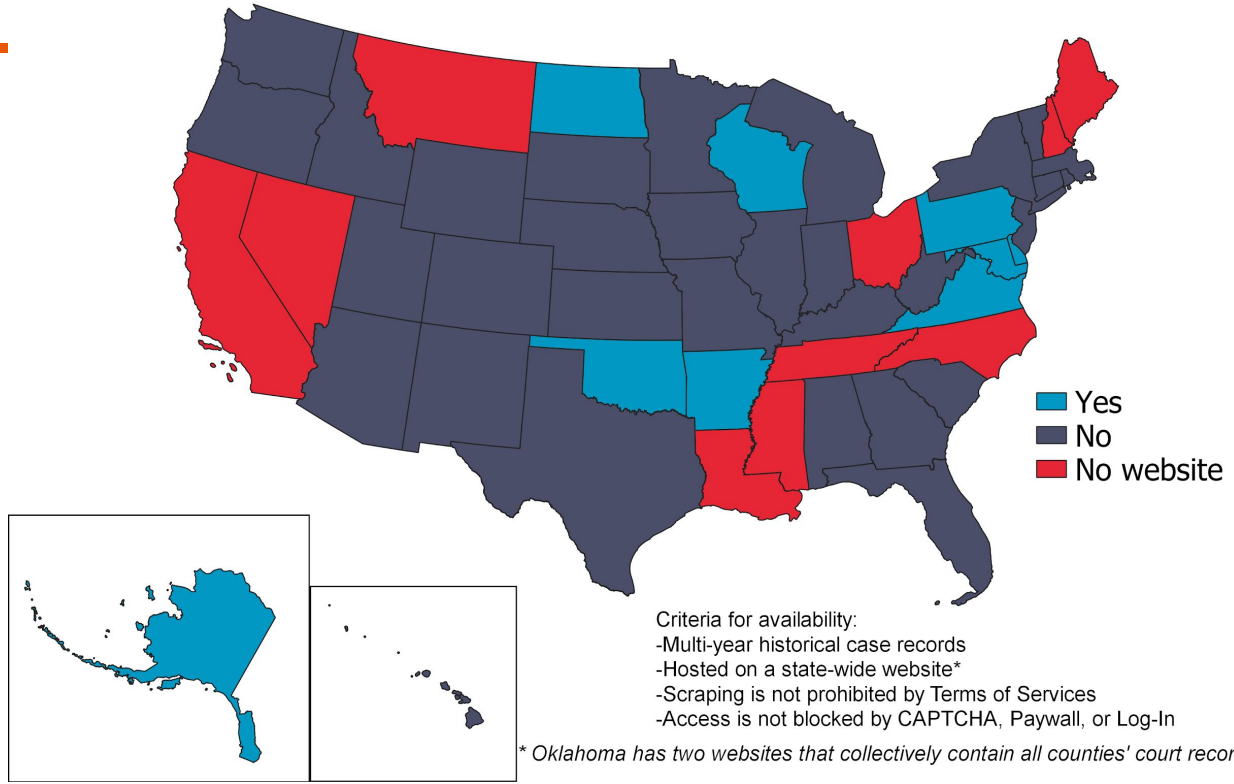
OTHER EVENTS AND HEARINGS

01/03/2000 **Case Filing**
CASE FILING | CASE FILING
01/03/2000 **Payment Receipt**
RECEIPT OF PAYMENT | PAYMENT RECEIVED
01/03/2000 **Paper**
ISSUE PAPER
01/03/2000 **Attorney Appointed to Defn Pursuant to SSRA**
FORC | FORCIBLE DETAINER CITATION
01/03/2000 **Historical Comments**

Tarrant County, TX Justice of the Peace Court.

So what's the problem?

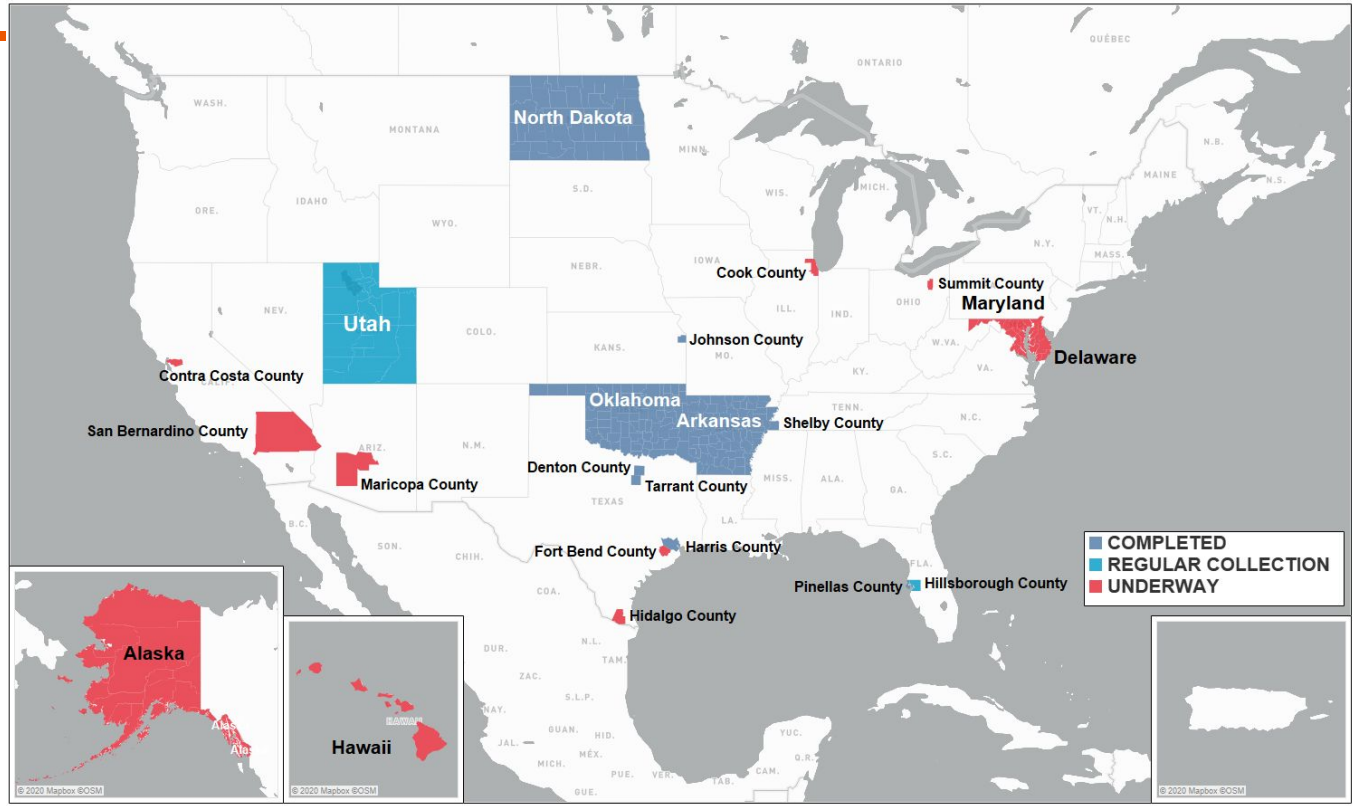
Court Data Availability via Statewide Portals, 2020



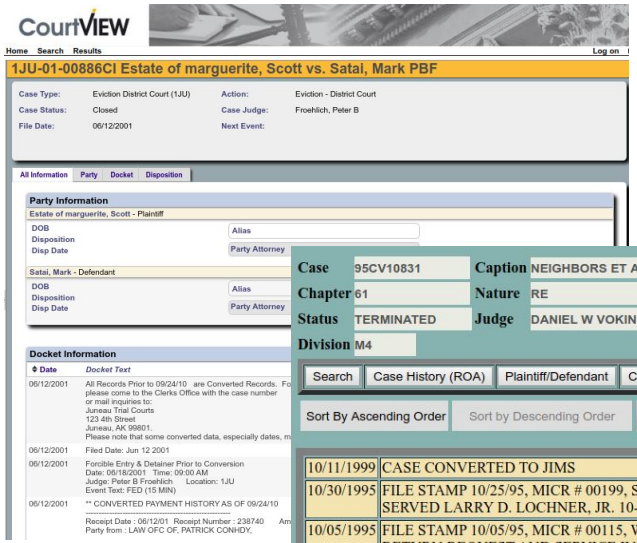
Criteria for availability:
-Multi-year historical case records
-Hosted on a state-wide website*
-Scraping is not prohibited by Terms of Services
-Access is not blocked by CAPTCHA, Paywall, or Log-In

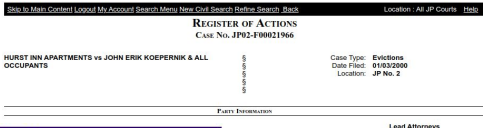
** Oklahoma has two websites that collectively contain all counties' court records.*

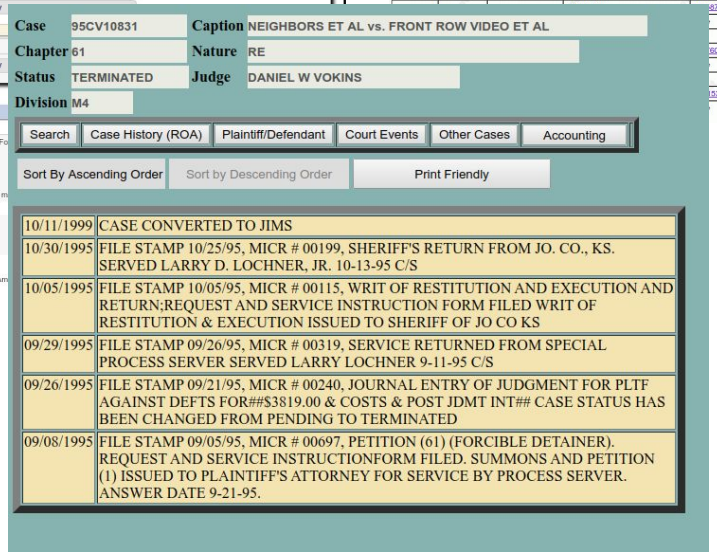
Data Collection (April 2020)



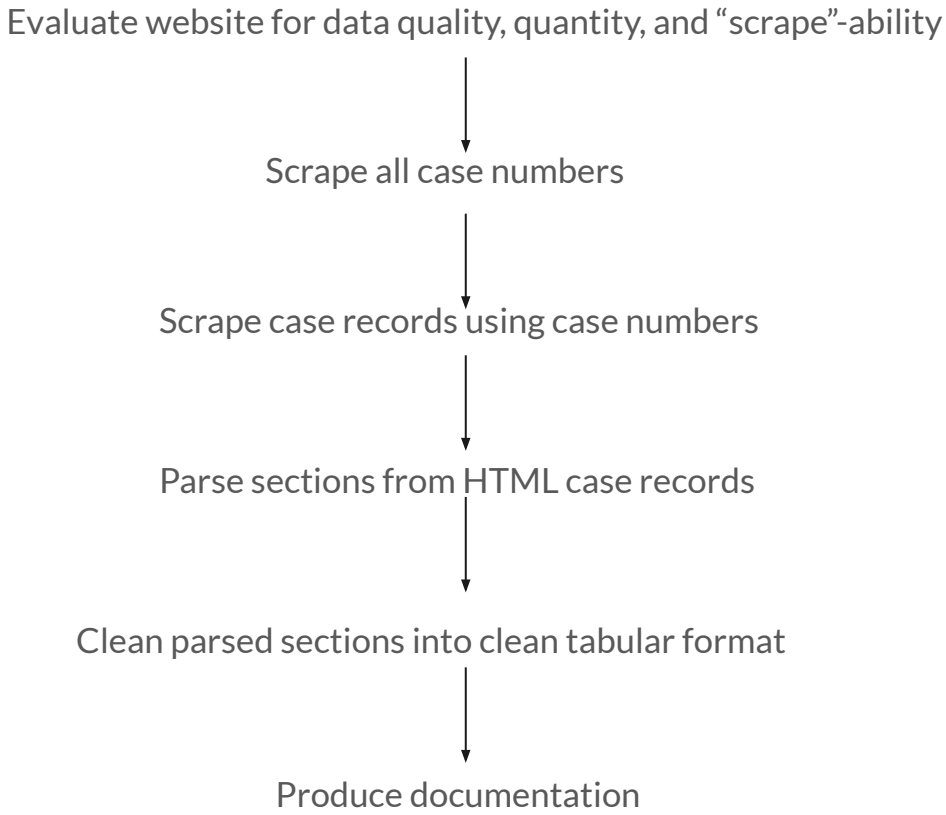
Court data comes in all shapes and sizes



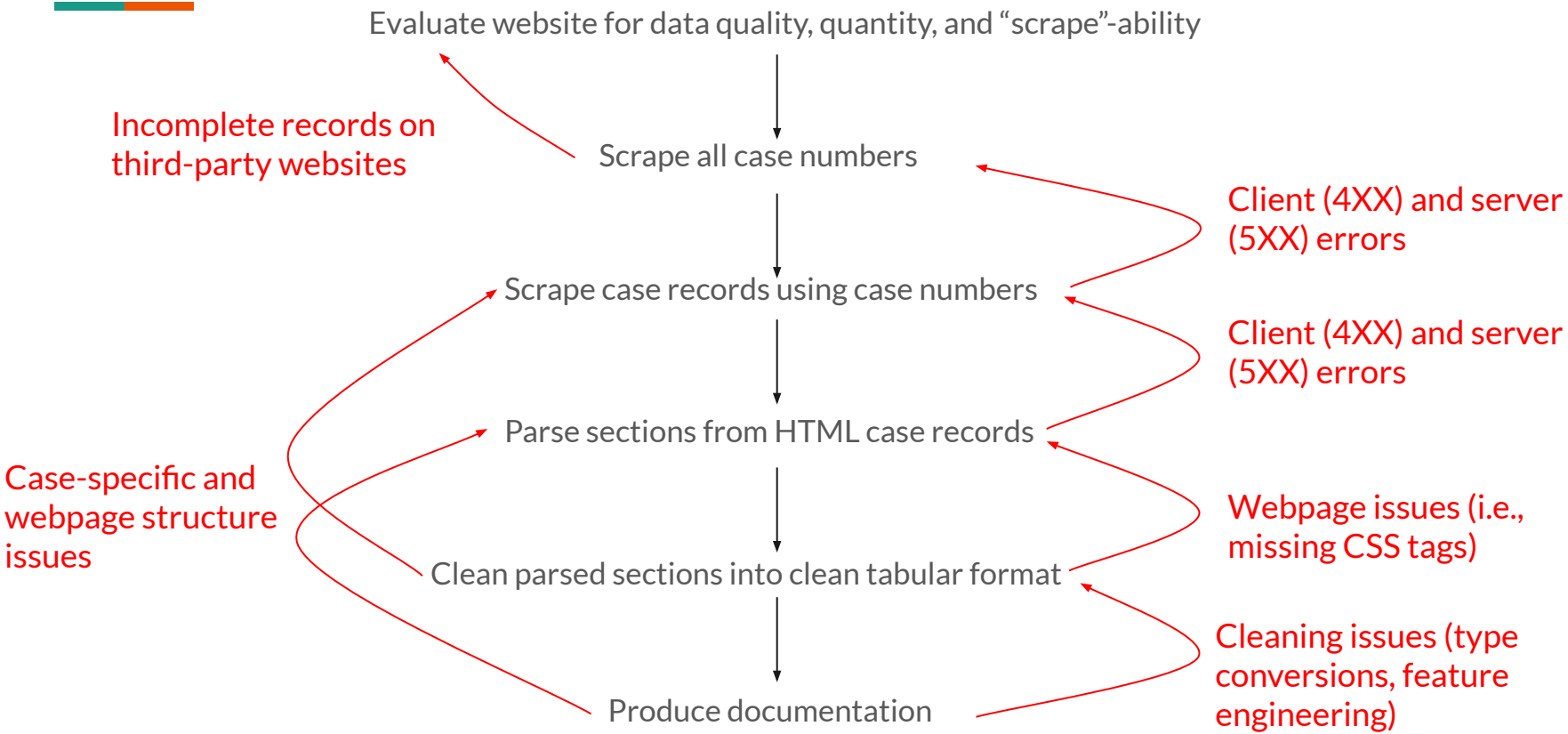




What I think my workflow will be...



What my workflow actually is...



Lessons Learned: Web scraping

- Be respectful
  - Monitor website response times and adjust your requests accordingly.
  - Plug: Urban Institute's SiteMonitor makes this easy and pain-free
- Be resilient
  - Find ways to "break" the website to understand error patterns: try big queries, small queries, clearly erroneous queries
  - Add explicit verification steps that confirm the expected inputs and outputs of scraping, parsing, and cleaning
- Be resourceful
  - Understand the network activity required for accessing web data.
  - Plug: Chrome DevTools provides methods for inspecting

Lessons Learned: Creating datasets

- Be clear about the intention of your final dataset
  - Do you strive to preserve the structure of the raw data?
  - Do you need to standardize the data model across different sources?
- Understand how users will access the data
  - Will there be live connections to dashboards?
  - If data will need to be updated, what data formats makes this easy?

I want to hear from you!

For questions, collaboration, and data sharing, please reach out:

Dan Bernstein

bernsteind@lsc.gov

@danbernstein18