

The Challenge of Big Data Legal Analytics Software –

How Dangerous are Judicial Profiling Services, and Is a Ban Really the Right Way to Deal with Them?

Zsolt Zódi PhD, habil, senior research fellow

Institute for Information Society
University of Public Service, Hungary
Digital Justice Research Project

*New Technologies in Courts:
Advantages and Limits*

Sapientia Hungarian University of Transylvania
Department of Legal Science

The story

The Ban

- In France, in March 2019, as part of the judicial reform (LOI n ° 2019-222 du 23 mars 2019 de programmation 2018-2022 et de réforme pour la justice), the Judicial Administration Act was amended and French court judgments were made fully accessible to the public. (This is notable Article 33)
- At the same time, however, two new rules have been added to the legal system
 - According to the first, the data of court officials and judges cannot be used to evaluate, compare and forecast their professional work (pratiques professionnelles).
 - Violators of the ban are the French Criminal Code. 226-18 (misuse of personal data).
- Explanation (Langford - Madsen, Verfassungsblog): the publicity of the huge disparities in refugee cases triggered this reaction - a man named Benesty published a huge sample of research on the Supra Legem site which found some judges reject almost all applications, while others give up almost everything.
- The judges immediately reacted hysterically and sided with the press: Supra Legem was the “aggressive Uber” of legal information, and so on. The provision may be part of this.

Reactions

Con

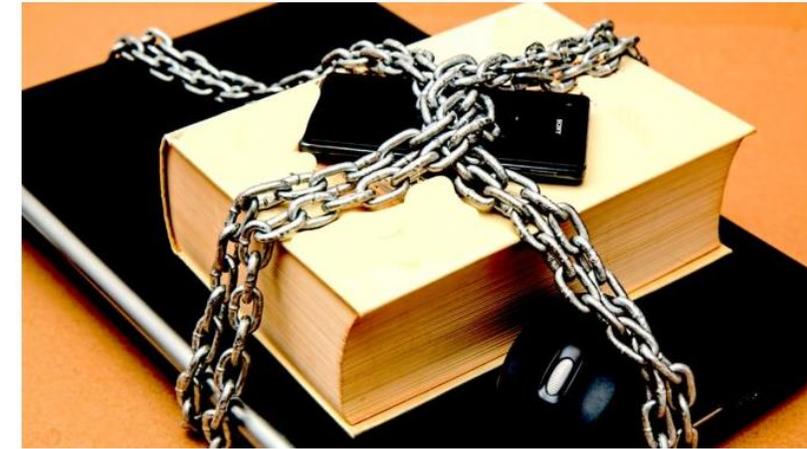
- Artificiallawyer.com: This is the first ban of its kind.
- Common law judges have already accepted it, but it is unusual for French judges
- At first, they also wanted to anonymize the names of the judges!
- In the end, this did not happen, but name-centric statistical processing was banned
- „This new law is a complete shame for our democracy”

Pro

- Prager v Austria (ECtHR): There are limits to the ability of judges to be judged. Cf. also Daily Mail: "Enemies of the People" front page with a picture of High Court judges after the first Brexit decision
- French judges judge “on behalf of the republic” and usually in a tripartite chamber, their judgment taking the form of syllogism, (“we did not decide, it follows from the law”)Although it is not possible to compile statistics on individual judges, it can be from “Council Z of Case XY Court 3”.All other analysis and statistics are still possible

France Bans Judge Analytics, 5 Years In Prison For Rule Breakers

4th June 2019 artificiallawyer Litigation Prediction 45



The system

Background

- In the judiciary and government, more and more digital data is being created and made public: it attracts analysis and processing.
- Scientists have long performed various statistical analyzes on such data.
- New problem (and opportunity): the machine can find relationships in seconds that would have to wait weeks for manual processing, or it would be completely impossible to create - e.g. interconnected networks of companies, companies controlled by one person in a company database. You are on the border of the legal gray zone now too!
- There is a fundamental difference between the US and Europe in two respects: in the US, the protection of personal data (which is not even so called) can be overridden by a number of principles, e.g. considerations relating to the publicity of court proceedings.
- In the U.S., not only court judgments but also (at the federal level) documents are largely public.
- “Legal analytics” as a genre originated in the US

WHO CAN ACCESS PACER?

PACER is available to anyone who registers for an account.

WHAT INFORMATION IS AVAILABLE ON PACER?

PACER hosts millions of case file documents and docket information for all district, bankruptcy, and appellate courts. These are available immediately after they have been electronically filed. If you are unable to locate the information you are seeking, please contact the local court or the PACER Service Center at (800) 676-6856 between the hours of 8 AM and 6 PM CT Monday through Friday or by email at pacer@psc.uscourts.gov.

IS ALL CASE INFORMATION AVAILABLE TO THE PUBLIC?

No. Some case information is protected. Certain personal identifiers are removed or redacted before the record becomes public, including Social Security number, financial account numbers, the name of a minor, date of birth, and home addresses in a criminal case. In addition, some documents may be unavailable to the general public, including:

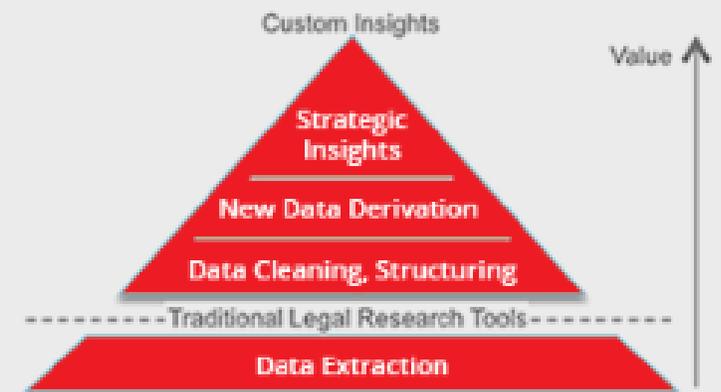
- Pre-2003 bankruptcy case documents.

The philosophy of analytics

- They go beyond traditional legal source research tools because more,
 - with cleaner data
 - They provide analyzes from a variety of perspectives,
 - are able to predict,
 - they provide information not only on the sources of law but also on other marginal circumstances of the case.
- „Beyond search, analysis"

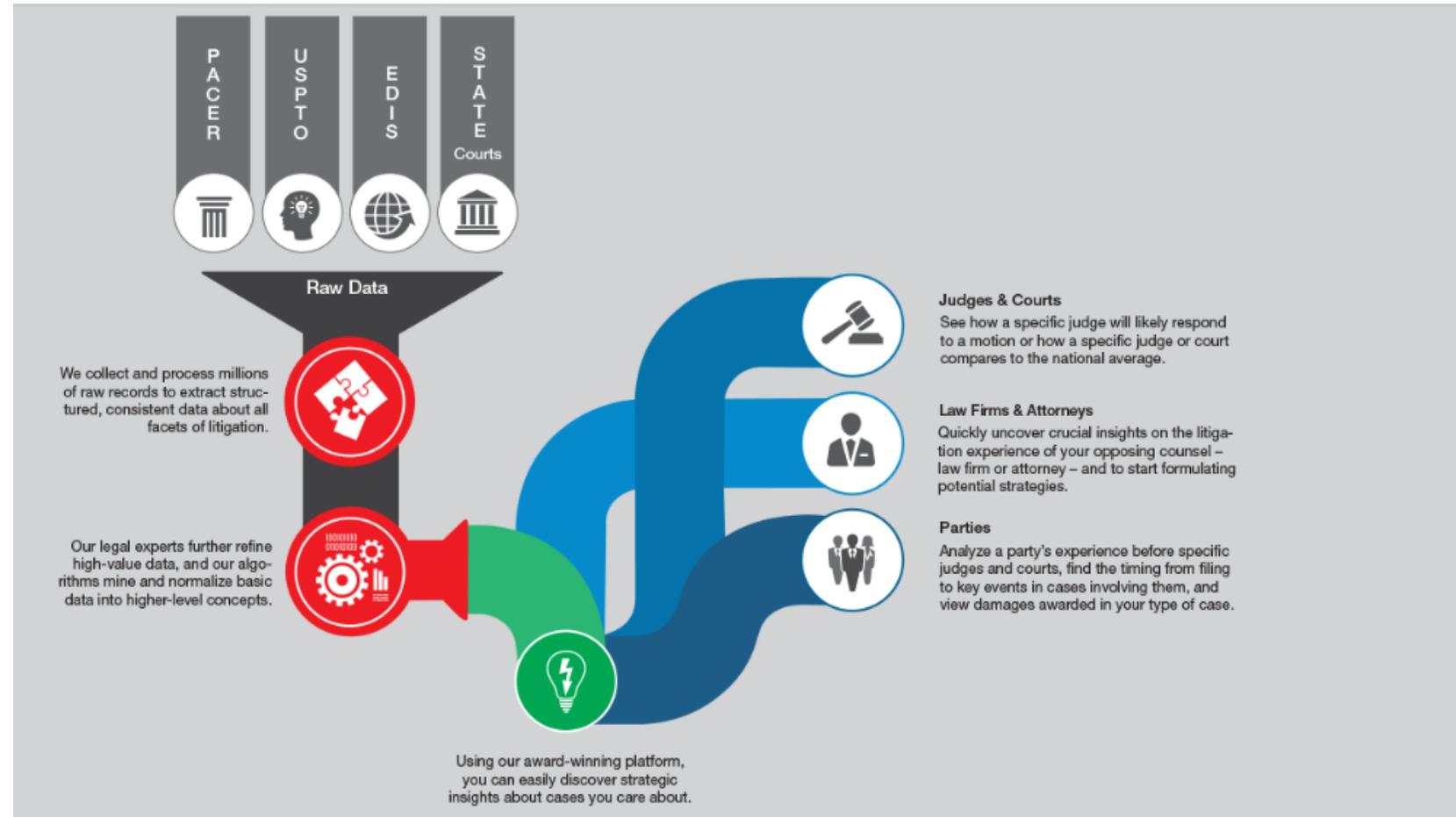
The Insight Gap

Traditional legal research tools focus on simple data extraction, but they don't do the difficult work to clean and structure their data. Limited to answering legal research questions, traditional tools end up presenting you with a list of cases that makes it hard to glean insightful information, unless you drill into each and every case. Lex Machina goes further—much further.



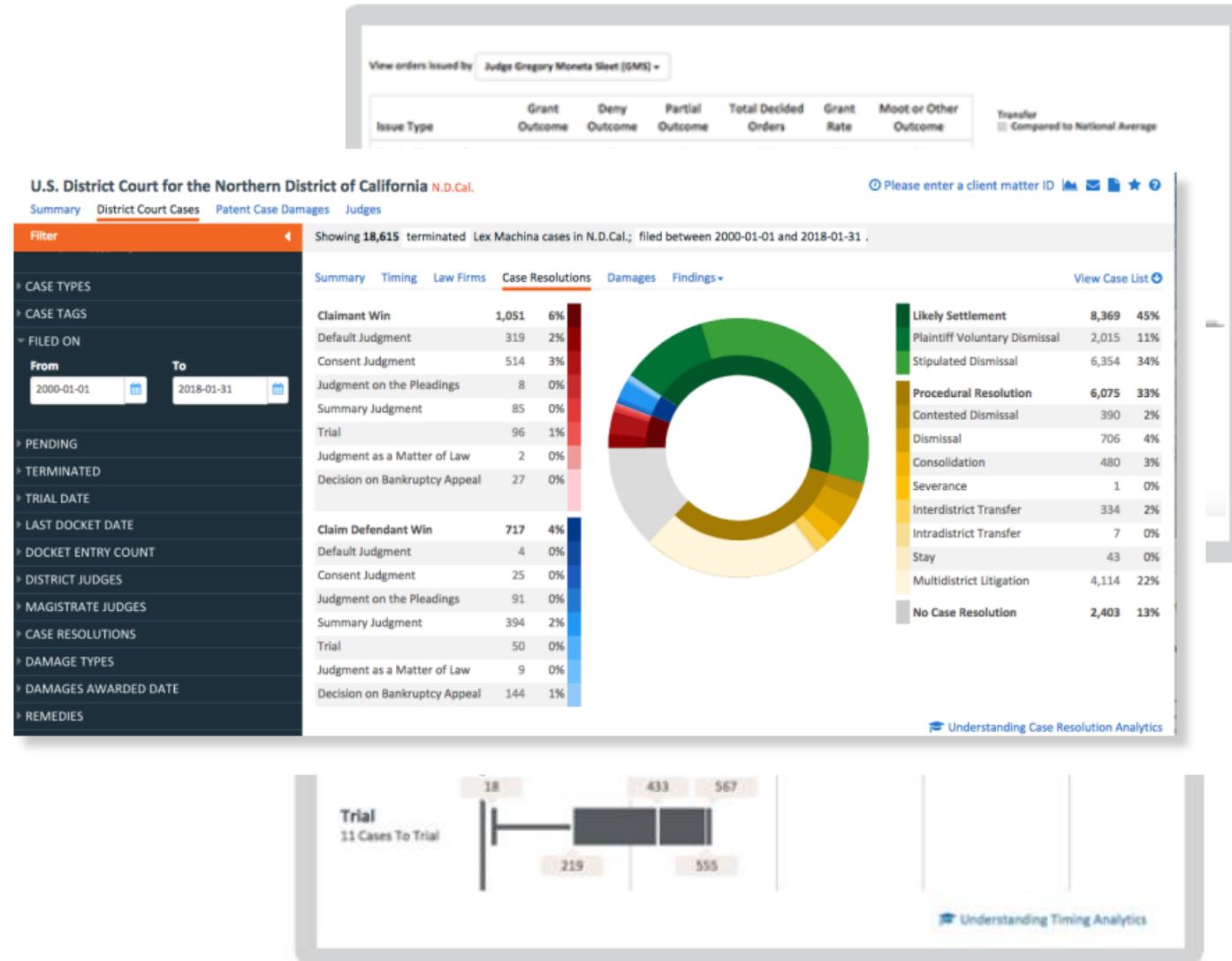
Technical background of the analysis

- Three phases:
 - Machine data extraction
 - Manual editing, data enrichment
 - Preparation of machine analyzes, publishing in various forms



Types of analytics

- Data can be collected and displayed in multiple views
- Statistics in tabular and graphical form
- Participants of the parties
- Judges and courts
- Case types,
- Precedents (list and network)
- Damages awarded
- Length of the procedures and procedural steps



Judge analyzer

- Judge's portfolio (case type per judge)
- Lawyers and their success rate before the judge (attorneys before judge - win ratio)
- Length of cases
- Precedents favored by the judge
- Popular arguments, phrases
- In whose favor do you typically judge and in what matters?

Case Type Per Judge

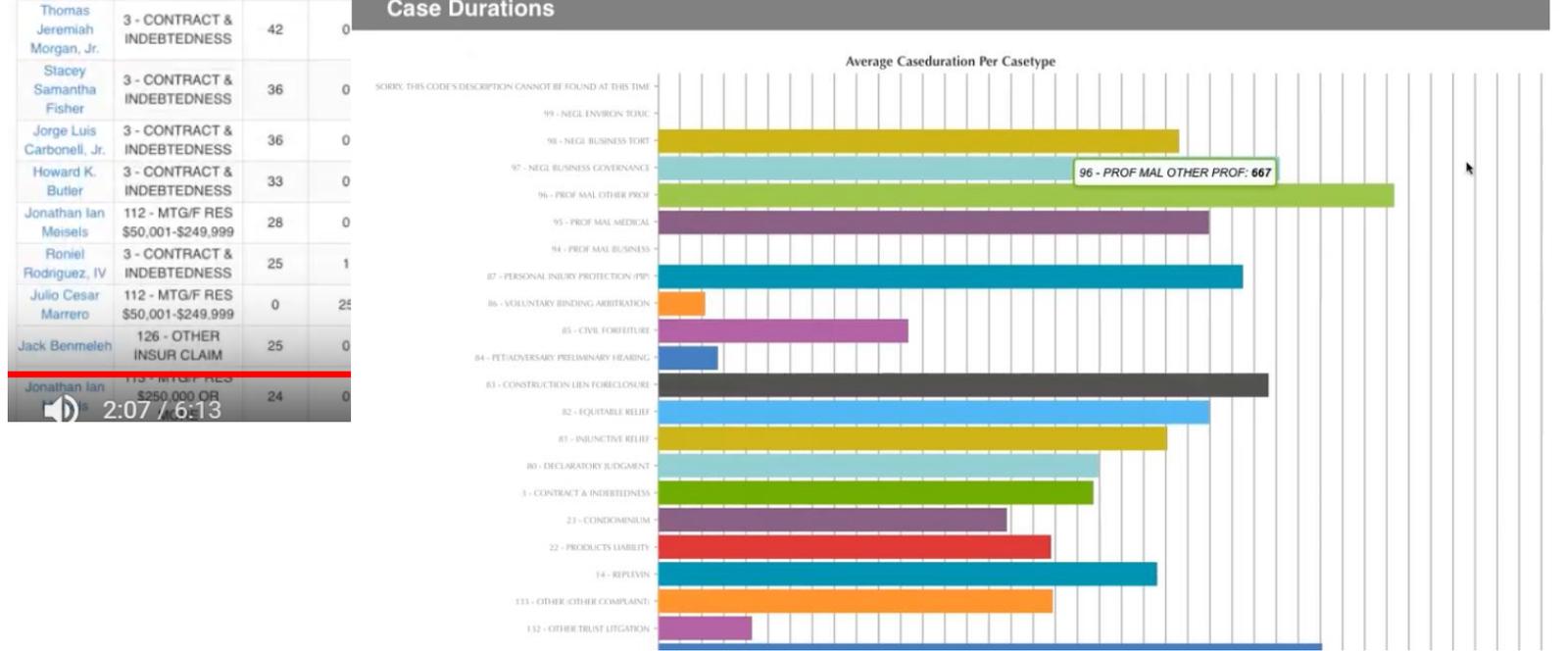
Casetype	Total No Of Cases	% Of Closed Cases	Average Caseduraton	Plaintiff Outcome
1 - EMINENT DOMAIN	8	87.5	532.43	42.86 %
10 - AUTO NEGLIGENCE	558	84.23	452.85	5.11 %
100 - NEGL THRID PART INDEM	2	100	798	0
101 - NEGL CONSTRUCTION DEFECT	13	76.92	323.5	10 %
102 - NEGL MASS TORT	1	0	0	0

Attorneys Before Judge

Megnézendó videóok Megosztás Inform

No of records per page Select

Attorney	Casetype	Plaintiff Attorney	Defendent Attorney	Open Cases	Winning Count	Losing Count	Total Count	Closed Plaintiff Cases With Dispositions	Plaintiff Win Count	Plaintiff Outcome	Closed Defendent Cases With Dispositions	Defendant Win Count	Defendent Outcome	Winning Percentage	Losing Percentage	Overall Outcome	Avg Case Duration
Kenneth Robert Duboff	126 - OTHER INSUR CLAIM	86	0	9	2	73	86	77	2	2.6 %	0	0	0%	3 %	95 %	2.6 %	529.13
Anthony Manuel Lopez	3 - CONTRACT & INDEBTEDNESS	57	0	11	0	46	57	46	0	0 %	0	0	0%	0 %	100 %	0 %	401.8
Philip Anthony Orsi	3 - CONTRACT & INDEBTEDNESS	56	0	0	16	40	56	56	16	28.57 %	0	0	0%	29 %	71 %	28.57 %	342.73
Scott Strems	3 - CONTRACT & INDEBTEDNESS	54	1	10	0	44	55	44	0	0 %	1	0	0%	0 %	98 %	0 %	368.67
Leo A. Manzanilla	126 - OTHER INSUR CLAIM	48	0	4	0	44	48	44	0	0 %	0	0	0%	0 %	100 %	0 %	489.55
Thomas Jeremiah Morgan, Jr.	3 - CONTRACT & INDEBTEDNESS	42	0														
Stacey Samantha Fisher	3 - CONTRACT & INDEBTEDNESS	36	0														
Jorge Luis Carbonell, Jr.	3 - CONTRACT & INDEBTEDNESS	36	0														
Howard K. Butler	3 - CONTRACT & INDEBTEDNESS	33	0														
Jonathan Ian Meisels	112 - MTG/F RES \$50,001-\$249,999	28	0														
Roniel Rodriguez, IV	3 - CONTRACT & INDEBTEDNESS	25	1														
Julio Cesar Marrero	112 - MTG/F RES \$50,001-\$249,999	0	25														
Jack Benmeleh	126 - OTHER INSUR CLAIM	25	0														
Jonathan Ian Meisels	112 - MTG/F RES \$250,000 OR	24	0														



So, what to do now?

Is ban a good way forward?

- With a ban, we are definitely not going anywhere. The legal sphere needs to be prepared for these analyzes.
- Already where there is data! In most EU countries there is NOT ENOUGH DATA for analysis and what is not available due to (other) legal barriers
- Hungary - litigation output: approx. 400,000 judgments, published: 10,000. There is no prospect of lower court judgments. No access to documents other than judgements. (They are now at least digitally available ...)
- Privacy concerns - information about litigants - even if they are not natural persons - cannot be extracted
- This is sad in two ways:
 - We need to lack an innovative service (cf. knowledge industry, international competitiveness - EU vs China vs US competition for priority)
 - Data-based analyzes could be an irreplaceable source for jurisprudence and legislation, but in the long run they could also improve the quality of law enforcement.