IOWA

PETS 2023

# GDPRxiv: Establishing the State of the Art in GDPR Enforcement

**Chen Sun**

*with Evan Jacobs, Daniel Lehmann, Andrew Crouse, and Supreeth Shastri*

https://GDPRxiv.org

# Motivation

**§17: Right To Be Forgotten** Right to obtain the erasure of personal data without undue delay

However, from a computing perspective, complying with RTBF leads to many **uncertainties**

### Deletion Latency
How soon should the data be deleted after the request?

### Deletion Depth
Delete data at the application level or all the way thru' hardware?

### Deletion Granularity
Should you allow deletion by individual item, by category, or all-or-nothing?

### Deletion Propagation
Should you inform other controllers or processors?

# How to reduce uncertainty in complying with GDPR?

**Track** GDPR enforcement in the real-world,
and then **adapt** the computing systems to meet/exceed the observed standards

## State of the Art (SOTA) in GDPR

A set of technologies, designs, mechanisms, policies, configurations, and operational practices that have failed to pass the current legal standards of GDPR compliance

### Research Goals

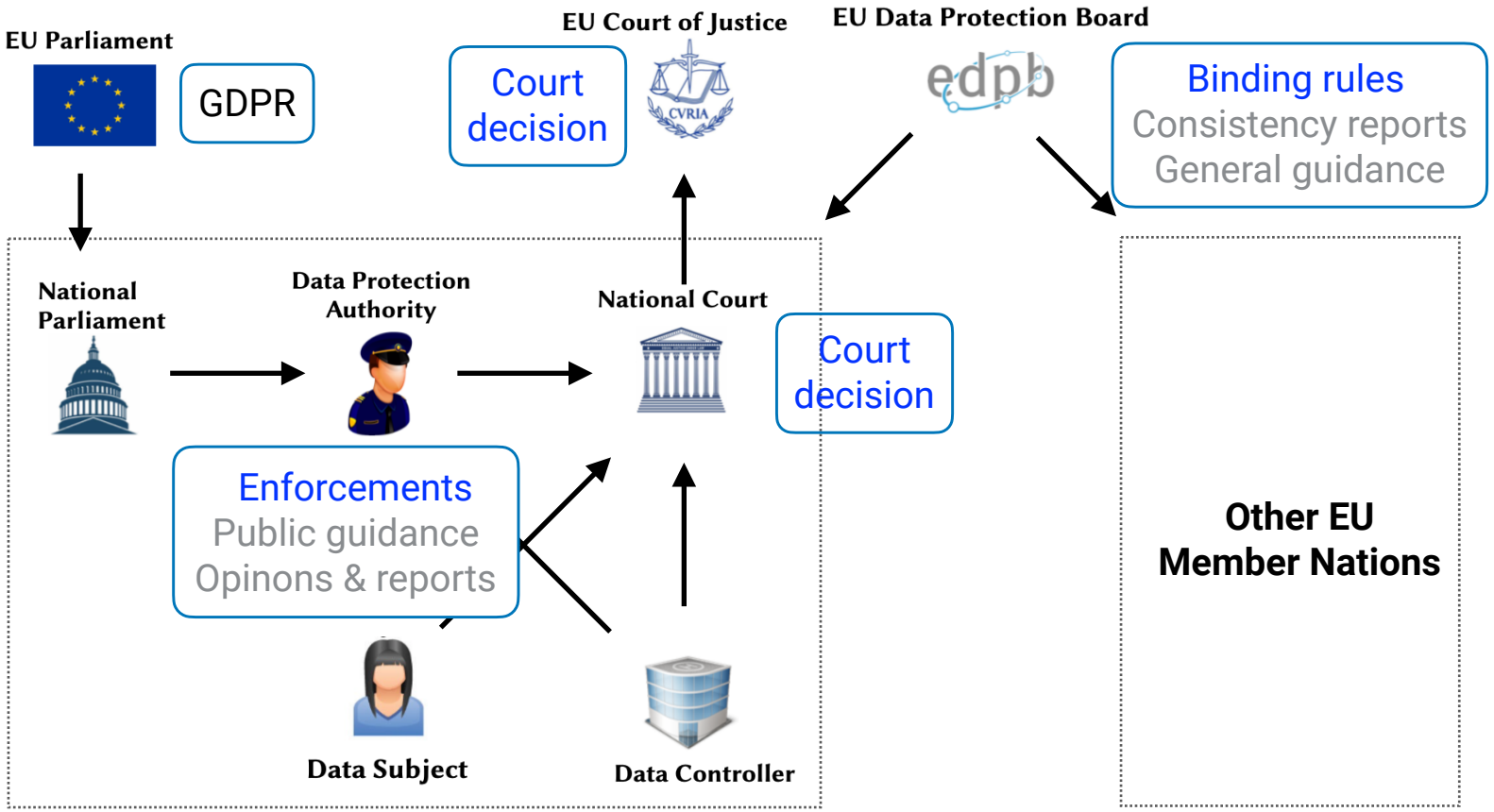
- Understand and model how GDPR enforcement works
- Build an open-source system to automatically capture GDPR SOTA documents
- Demonstrate the utility of the system

# SOTA Documents and How to Get Them

1. Identifying source of information

2. Characterizing the Information
   - ➡ Legal precedent
   - ➡ Legal guidance

3. Procuring the information
   - ➡ GDPR-aware crawler



# Architecture

**Official Sources**
DPAs, EDPB, courts, EU parliament

*A total of 47 websites consist of all the SOTA-defining documents*

**terface**
re corpora; or explore
untry, articles, penalty etc

**Crawler Engine**
HTML Parser, Policy Mgmt

**SOTA DB**
all SOTA knowledge

*8943 documents in 4.5 years*

**Data Curator**
manage GDPR raw data

**Label Engine**
identify key characteristics of documents

1. Filter out non-GDPR documents
2. Classify document by types
3. Translate to English

Key metadata

- Country, language, document type, release date (automatically labeled)
- Decision type, controller, cited GDPR articles, financial penalty (manually labeled)

## GDPRxiv

- the first ***open source*** GDPR crawler; and an ***open access*** SOTA knowledge base
- consists of **2X** more legal precedents, and **5X** more SOTA documents than prior work

# How Can GDPRxiv Help?

> *Track the enforcement of **GDPR over the years***
>
> *Help reduce compliance uncertainties using the **via negativa analysis***
>
> . . .

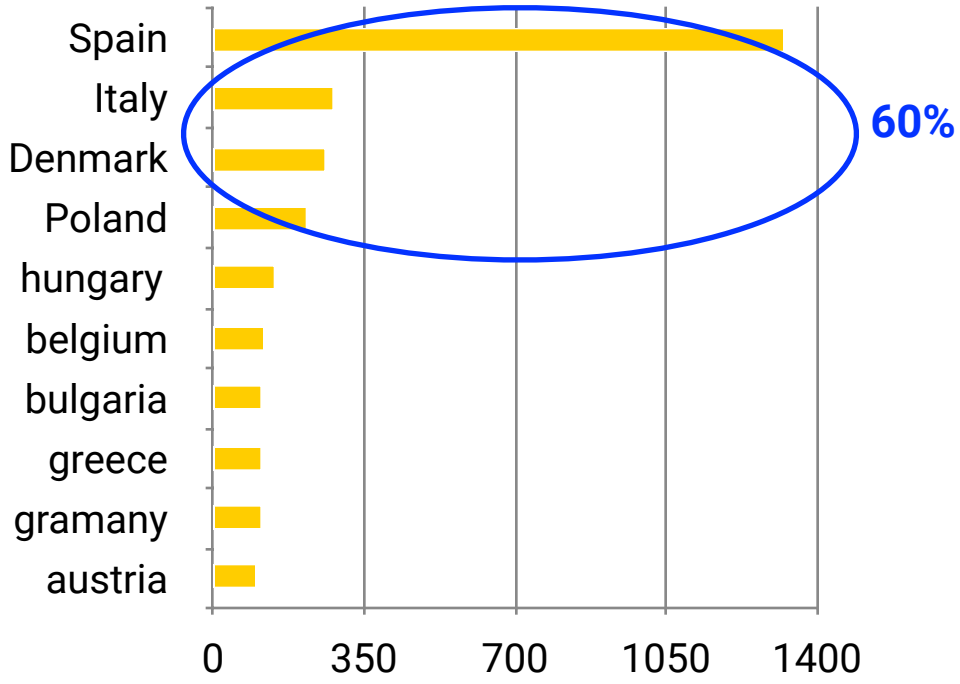
# Enforcement Trends

### GDPR is not implemented uniformly across Europe

- Four countries (Spain, Italy, Denmark, Poland) account for ***60%*** of all GDPR enforcements

### Enforcements are issued frequently and growing over time

- On average, **2 enforcement decisions** are issued every day
- Year 4 saw **2.7×** more enforcements than year 1

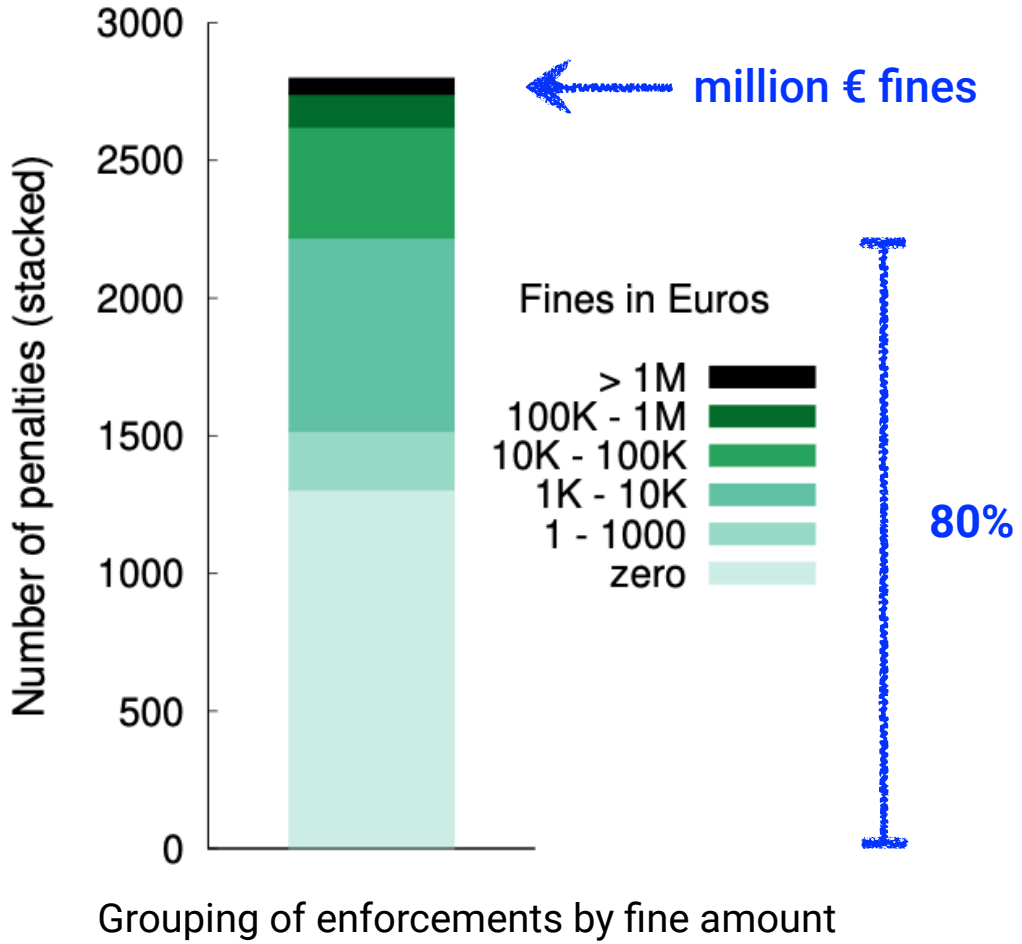


Top ten countries w/ the most GDPR enforcements

# Enforcement Trends

## € Practice of proportional penalties

- **80%** of the fines were less than €10K
- Only **1.8%** violators ended up with Million€ fines



# Reducing uncertainty in GDPR compliance : *Via Negativa approach*

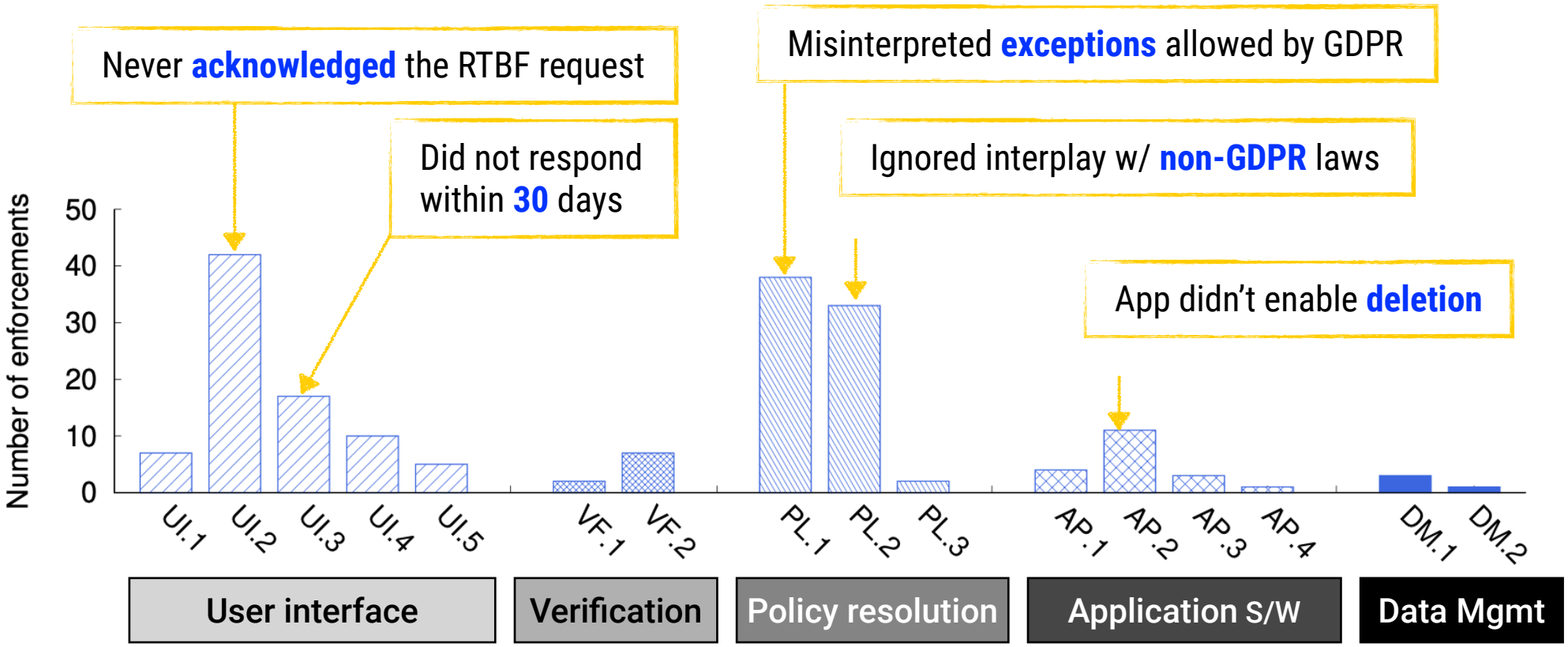

Laws including GDPR tends to be *under-specified* and *all encompassing* *from a computing perspective*


Our SOTA allows us to identify, definitely, how organizations have failed to comply with GDPR.

Via Negativa approach: ***eliminate these failures*** to reduce your risk and uncertainties

# Via Negativa analysis on RTBF

> *We analyze all enforcements that cite article-17 RTBF (= 175 in the first 4.5 years);*
> *then, we extract the main reason(s) that led to the enforcement*




**80%**

failures are in UI or Policy

# Summary

## Uncertainty in complying w/ GDPR

Translating legal intentions to computing implementation is fraught with uncertainties

## GDPR State of the art

We propose a method to establish a reliable source of ground truth in GDPR enforcement

## GDPRxiv

Open-source crawler and knowledge base;
Largest ever collection of GDPR enforcements;
Novel insights derived from real enforcements

## How can GDPRxiv help you?

Please check out **https://GDPRxiv.org**
We love to hear about your use cases and suggestions for improvement