GDPRxiv: Tracking GDPR Enforcement in the Wild

Daniel Lehmann
Computer Science
University of Copenhagen

Chen Sun
Computer Science
University of Iowa

Andrew Crouse
School of Law
University of Iowa

Supreeth Shastri
Computer Science
University of Iowa

ABSTRACT

Though European Union's General Data Protection Regulation (GDPR) is hailed as a model privacy regulation, details about its enforcement are not well understood. To address this gap, we propose establishing the state of the art (SOTA) in GDPR enforcement, and present the design and implementation of *GDPRxiv*: an information archival system that collects and curates GDPR rulings, judgements, reports, and official guidances. *GDPRxiv* consists of the largest centralized collection of GDPR knowledge base and helps us identify novel insights from the first three years of GDPR enforcement.

Artifact Availability:
The source code, data, and/or other artifacts have been made available at https://GDPRxiv.org.

1 INTRODUCTION

> *"One of the great mistakes is to judge policies and programs by their intentions rather than their results."*
>
> Milton Friedman (1975)

The General Data Protection Regulation (GDPR) [25] has been in effect since May 2018. It was the first major law to elevate the privacy and protection of personal data to be a fundamental right, and then accord that right to 450 million people of Europe. Since then, GDPR has emerged as a model regulation for data protection efforts around the world [2, 4, 6, 14]. Despite its outsized influence on data protection debates and policies around the world, details of its enforcement are not well understood. For example, there is no comprehensive repository of all the GDPR rulings, judgements, advisories, reports, and guidances; nor have there been any systematic analysis of its enforcement trends; instead, much of the focus has been on big monetary penalties levied on popular companies.

Absence of such comprehensive ground truth has rendered compliance efforts to be ad hoc and narrative-based, which further jeopardizes the protection of data and exposes organizations to legal risks. We illustrate how this uncertainty in interpreting GDPR has manifested at every stage of the design and operation of computing systems (in Section-2.2). To alleviate this situation, we propose establishing *the state of the art (SOTA) in GDPR enforcement.* We define *GDPR SOTA* to be a set of technologies, designs, mechanisms, configurations, and operational practices that have failed to pass the current legal standards of GDPR compliance. Most scientific and legal disciplines require having a clear understanding of what the state of the art is at any given time. Thus, the goal of our work is to build such a knowledge base and make it available to the computing community.

While it is important to understand GDPR's enforcement holistically, it is challenging for two reasons. First, *the decentralized nature of its enforcement.* Though GDPR is legislated by a centralized entity, namely the EU parliament, its enforcement is handled by 30+ independent entities called Data Protection Authorities (roughly, one per EU country). This has led to considerable divergence in enforcement priorities, practices, and timelines across Europe. Second, *our collective understanding of data rights and responsibilities is still evolving.* Introducing a new right into the society is a long drawn out process, where stakeholders gradually converge towards an equilibrium point. Consider the journey our society has gone through for women's rights, voting rights, or civil rights; GDPR and data protection rights are just three years in the wild. Thus, any effort to establish GDPR SOTA must interface, *comprehensively and continually*, across all official sources of enforcement.

We begin our work by modeling how GDPR is legislated, enforced, and interpreted. This, in turn, helps us recognize the sources and characteristics of the enforcement information that constitute the ground truth. Then, we design and implement a GDPR-aware crawler that procures these data from official sources over the Internet. As a result of this effort, we have put together the largest centralized collection of GDPR enforcements, judgements, opinions, reports, and guidances. Finally, we build *GDPRxiv*[1], an information archival system to automate the collection and curation of enforcement data; to organize and analyze the procured legal corpora; and to disseminate the knowledge to the computing community.

Our analysis of GDPR enforcement corpora brings out several novel insights about enforcement activities, priority areas, targets, and financial penalties. Table-1 provides a concise summary of these findings. While three years is a short time to judge the efficacies of a transformative regulation like GDPR, our findings do reflect that its enforcements are broadly aligned with its original intent. Our long-term vision is to evolve GDPRxiv into a platform for data-driven research and analysis concerning GDPR compliance and enforcement.

Summary of contributions. Our work identifies and solves a foundational problem in the emerging area of privacy regulations. In particular, we make the following contributions:

- **GDPR SOTA:** We describe the need for and a means to compose the state of the art in GDPR enforcement. We model GDPR's implementation ecosystem in Europe towards identifying the key sources and the characteristics of enforcement information.
- **GDPRxiv:** We present the design and implementation of GDPRxiv, a GDPR-aware crawler and an archiver of its legal corpora. GDPRxiv, to our knowledge, is the first and only system to be completely automated, be open-sourced, and to expand on the previously existing GDPR corpora by 7×. We

[1] GDPRxiv is a portmanteau of GDPR and arXiv, pronounced as G-D-P-archive

Table 1: ***Key findings (in blue) and high-level insights (in black) from our analysis***

Enforcement activities	GDPR is not implemented uniformly across Europe. *Five DPAs account for 49% of all GDPR activities.*
	GDPR activities are growing over time. *Year-3 generated 50% more SOTA-defining documents than year-1.*
Financial penalties	Fines were used frequently but there is a heavy skew in their application. *On average, a GDPR fine was issued once every 1.4 days.* *399 (or 52%) of the fines were for less than €10K.* *Top 10 highest fines accounted for 60% of all the fine amount.*
Focus areas	Regulators are prioritizing sound and secure practices of data management over reports of data breaches or failures to honor an individual's rights. *Three articles (5, 6, 32) account for 65% of all fines and 59% of all citations.*
Enforcement targets	Large companies have borne the brunt due to proportional penalties. *Large companies received ∼same number of enforcements as small/micro companies yet they paid 94% of total amount of fines.*

will publicly release all our software artifacts and datasets at https://www.GDPRxiv.org.

- **Insights and Opportunities:** We share novel insights gleaned from a systematic analysis of the enforcement corpora. We also lay out our vision for GDPRxiv as a data platform for research and education and as a bridge between law and CS communities.

2 BACKGROUND AND MOTIVATION

In this section, we discuss the importance and timeliness of the problem, review related work, and establish the need for and novelty of our work.

2.1 GDPR and its Emergence as a Model Regulation for Data Protection

GDPR [25] is a European regulation that declares the privacy and protection of personal data to be a fundamental right of all the European people, and assigns explicit responsibilities to companies that collect and process such personal data. It became enforceable in all EU member states from May 2018. A prominent feature of GDPR is that it allows regulators to impose hefty penalties (for instance, fines of up to €20M or up to 4% of the annual worldwide revenue, whichever is higher) on organizations failing to comply with GDPR.

Since GDPR was the first comprehensive data regulation and its initial roll out was effective, policy makers around the world began adopting GDPR as a template for their laws. For example, both California's Consumer Privacy Act (CCPA) [2] and Virgina's Consumer Data Protection Act (CDPA) [6] retain a majority of the core rights and responsibilities outlined in GDPR. This influence is not restricted to public domain alone. Microsoft, for example, has announced [14] that it would voluntarily extend the core rights of GDPR across the world. Legal scholars refer to this phenomenon as *the Brussel's effect* [13], a race to the top effect where the early but stringent standards of a EU regulation gets proactively applied beyond its intended geographical boundaries. Thus, given the foundational role of GDPR on other data regulations, it is imperative to understand GDPR's implementation in the wild.

2.2 Scale and Scope of Uncertainty in Complying with GDPR

We are in the early days of data protection regulations, where stakeholders (namely, companies, people, policy makers, journalists, CS/law scholars etc) are engaged in a tussle to define, adapt, and enforce data rights. We think of this period as what the 1920s were for the women's rights or 1960s were for the civil rights. Thus, when regulations are enacted, policy makers and legal scholars tend to limit their expositions to core legal principles that are broadly interpretable and will hold the test of time, instead of getting into the specific implementations of the current time. While legally prudent, this strategy invariably leads to uncertainties from a computing perspective. Below, we illustrate how these manifest at different stages of the design and operation of computing systems:

Example-1: Uncertainty at organization level. *GDPR[2], via $\mathscr{G}$ 5(1) (B) Purpose Limitation, mandates that personal data can only be collected and used for specific purposes.* This is a major departure from 50 years of computing evolution, where the notion of purpose has been associated with programs and models, while data is viewed as a helper resource that simply serves these high-level entities in accomplishing their goals. This portrayal of data as an inert entity has allowed it to be used freely and fungibly across various systems. In the post-GDPR world, when the French data protection commission saw that Google was collecting user's personal data in one system (Android OS) and using it to serve personalized ads in other services (like YouTube and Search), it fined [28] Google €50M for lacking legal basis for such purpose bundling. If we take

[2] henceforth, we will prefix GDPR articles with $\mathscr{G}$

this purpose limitation to the other extreme, where every piece of data from every person needs to have a specific purpose for every service, prior work [30] shows that it leads to significant storage overheads and performance slowdowns, on top of cumbersome user interactions. In between these two extremes, there exists a number of configurations that allow different tradeoffs in compliance risk vs. computing performance that organizations have to now choose.

Example-2: Uncertainty at design level. *GDPR, via 𝒢17 Right to be Forgotten, grants people the right to request deletion of their personal data and requires companies to abide by it without undue delay.* From a computing design perspective, this requirement is heavily underspecified. Consider the *latency of deletion* i.e., how soon after the request, should the data be removed. Designers could opt for a strict compliance by making deletions synchronously in real-time, or choose a relaxed compliance by allowing deletions to happen eventually. Prior work [30] has shown the effect of synchronous deletion on two popular database systems, Redis and PostgreSQL, both of which experienced a slowdown of up to 20%. On the other hand, eventual compliance allows stale data to linger in the system for unspecified amount of time, posing security and privacy risks. Second, consider the *depth of deletion* i.e. should the data be deleted from all memory and storage subsystems going all the way to hardware, or simply be forgotten at the service level. While taking the former approach leads to a strict form of compliance, it adds significantly to the latency and complexity of the deletion process. For example, Google cloud guarantees a through deletion of customer data from all their systems but requires up to 180 days to complete the operation [3]. There are many other design parameters to the deletion requirement as well as to the multitude of other GDPR requirements that amplify uncertainty at design level.

Example-3: Uncertainty at operations level. *GDPR, via 𝒢30 Records of Processing Activities and 𝒢33 Notification of Personal Data Breach, requires companies to monitor all accesses to personal data so that data breaches can be investigated and reported to affected parties in a timely manner.* For a system administrator supervising a personal-data store, this translates to creating an audit trail of all accesses to personal data. The language of the law allows a broad spectrum of configuration choices: at the strict end, this turns every read operation into a data-read followed by a log-write, which effectively reduces the database throughput by half. In fact, prior work [30] has shown that for realistic workloads such as YCSB [19], database performance drops by up to 5×. On the other hand, admins could set up relaxed compliance configurations such as (i) saving audit logs to the disk asynchronously, (ii) monitoring data accesses at random or predetermined intervals (say, logging every 100th operation), or (ii) omit monitoring altogether by relying on access-control-lists. While these options reduce performance overheads, they expose the administrator to the risk of missing unexpected realtime events. So, if and when a data breach happens, they would have no choice but to inform *all their customers* that *all of their data* may have been compromised. Thus, without knowing the current enforcement thresholds for personal-data monitoring, administrators cannot effectively analyze their risk-benefit tradeoffs.

2.3 Reducing Uncertainty by Tracking the Enforcement of GDPR

One way to reduce uncertainty in understanding and complying with GDPR is to track its enforcement in the real-world, and then adapt the computing systems to meet or exceed the observed standards. This would require following the legal precedent set via regulatory enforcements, court judgements, public guidance, and other information from official legal sources. However, doing so is challenging due to (i) GDPR's enforcement complexity and (ii) its evolving interpretation over time.

The first challenge stems from the distributed nature of GDPR implementation. While GDPR is written by a centralized entity, namely the European parliament, its enforcement is handed over to 30+ independent and distributed entities called the Data Protection Authorities or DPAs (approximately, one per European country). Though bound by the same underlying regulation, every DPA has the autonomy to determine its own priorities, develop its enforcement strategies, and has to operate within the budgetary resources allotted by its national government. This has led to considerable divergence in the way GDPR is enforced and implemented across the EU.

Second, GDPR enforcement is a constantly evolving phenomenon. While GDPR precisely defines its *legislative intentions* i.e., what it intends to accomplish in principle, it leaves to broad interpretations the *technical implementations* i.e., how a company should build and operate personal-data systems to meet its obligations as well as how a DPA should regulate the controllers (as detailed in Section-2.2). This disconnect is not accidental: introducing a new right into the society is a long drawn out process, where stakeholders gradually converge towards an equilibrium point. As GDPR goes through this journey, we expect its enforcements to constantly evolve and adapt based on feedback from the involved stakeholders.

Thus, any effort to track the enforcement of GDPR must interface, *comprehensively and continually*, across all the official sources. We describe two contemporary efforts[3] to track enforcement and discuss how their shortcomings undercut their utility as reliable sources of ground truth:

- **GDPR Enforcement Tracker** [24]: is a website and mobile app that displays penalties levied under GDPR. As of this writing, it consists of 607 entries covering data protection authorities from all EU nations. Its key shortcomings vis-a-vis our effort are: (i) keeping the data collection and analysis methods proprietary, and (ii) focusing only on cases where monetary penalties are involved.
- **GDPRhub** [29]: is a wiki-style information portal, populated by voluntary contributors, that provides commentaries on

[3] based on informal conversations, we are aware of some form of ground truth being curated by big tech companies but, these efforts are tailored to their business models and unlikely to be released publicly.

Table 2: ***Comparing GDPRxiv with contemporary efforts across five key metrics***

	Enforcement Tracker	**GDPRhub**	**GDPRxiv**
Collection method	Proprietary	Hand-curated by volunteers	Open-source crawler
Content types	DPA enforcements	DPA enforcements; Court judgements	DPA enforcements; Court judgements; Official opinions, reports, and guidance
No. of documents	808	943	7560
Interfaces	Website	Wiki; Newsletter	Website
Sustainability	Unknown	Needs person-hours proportional to the documents added	Fully automated

GDPR enforcement. As of this writing, it describes ~1000 decisions from courts and data protection agencies. The main issue with GDPRhub is that, just like Wikipedia, it cannot be considered a reliable source of ground truth since the quality and quantity of its content are governed by the skill level and availability of its voluntary contributors.

2.4 Research Goals

The goal of our work is to establish a reliable and comprehensive source of ground truth in GDPR enforcement. We begin by understanding how enforcements work in the GDPR ecosystem, identifying the responsible legal entities, and by characterizing the enforcement data produced by them (in Section-3). This modeling helps us define the state of the art (SOTA) in GDPR enforcement. To actually procure such data and compose a usable knowledge-base, we design and deploy two systems: (i) **GDPR Crawler**: a system for collecting and curating legal data concerning GDPR's implementation, and (ii) **SOTA Manager**: a system for organizing and and disseminating the GDPR SOTA knowledge. We refer to these two systems collectively as **GDPRxiv** (in Section-4). Table-2 summarizes the key differences between prior efforts and our work. Finally, we share insights and trends identified by our knowledge base in (Section-5).

3 STATE OF THE ART IN GDPR ENFORCEMENT

The notion of the *state of the art* (SOTA) is prevalent in both law and computing. For example, in patent law, SOTA is used to assess and assert novelty; in tort law, SOTA is invoked to establish the current standards of the profession; and in machine learning, SOTA represents the best of the results achieved by the ML models so far. We extend this notion to data protection regulations and define *GDPR SOTA* to be a set of technologies, designs, mechanisms, configurations, and operational practices that fail the current legal standards of GDPR compliance. In this section, we analyze GDPR's enforcement ecosystem in Europe with a goal to identify all the sources that generate enforcement information, and to propose a way to procure this information via automated means.

3.1 Identifying the Sources of Information

Figure-1 depicts a representation of the GDPR ecosystem. The flow of control starts at the European parliament that passed GDPR as a binding regulation on April 14, 2016 and made it enforceable from May 25, 2018. All the member nations of EU are required to adopt this regulation via their national parliaments, thereby setting up an agency responsible for overseeing the enforcement of GDPR within their national boundaries. These agencies, referred to as Data Protection Authorities or DPAs, serve as the single point of contact for people exercising their personal-data rights and for organizations needing to demonstrate GDPR compliance. Based on complaints from data subjects, reported data breaches, and any findings of irregularities, the DPAs investigate GDPR violations and issue penalties, warnings, notices, and other enforcement decisions. DPAs may also release public guidance on technologies, policy advisories and opinions, as well as annual reports.

While DPAs serve as the sole regulator for all GDPR matters, both data subjects and data controllers have the right to challenge the DPA decisions in judiciary bodies such as national courts and the EU Court of Justice. Finally, to ensure that the rules of GDPR are applied consistently across all the member nations, a trans-national agency called the European Data Protection Board (EDPB) has been set up [11]. In its role, EDPB issues consistency reports, binding rules, and general guidance for DPAs and data controllers. Thus, to get a holistic view of GDPR enforcements, we need to track information from the EU parliament, DPAs of all member states, the EDPB, national courts, and the European Court of Justice.

3.2 Characterizing the Information

We observe that two broad categories of legal content are generated: (i) *legal precedent*, which is a principle, practice, or rule that gets established following a DPA enforcement decision or a court judgement such that subsequent cases with similar situation will likely follow the previously established outcome, and (ii) *legal guidance*, which are recommendations, opinions, and reports issued by GDPR bodies to help stakeholders and to clarify compliance matters without being binding. Examples of precedent include court judgements, EDPB consistency rulings and binding decisions, and DPA enforcement decisions that have been vetted by the courts; while examples of guidance include EDPB legal guidance, DPA's

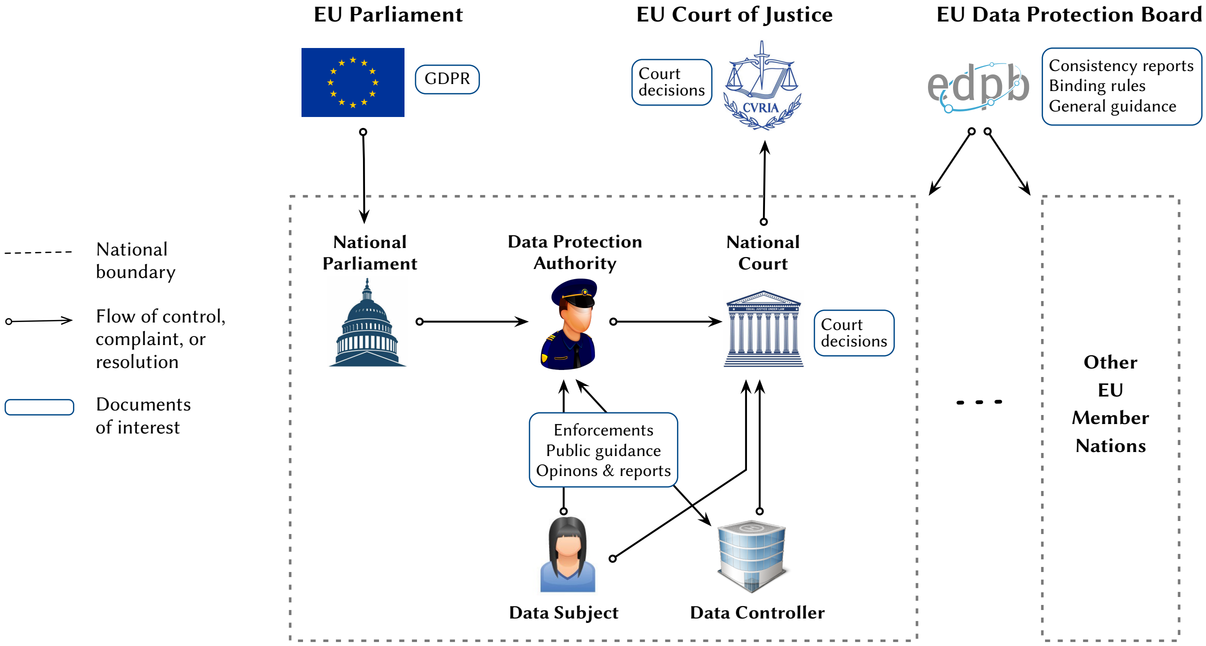


Figure 1: A representation of the GDPR enforcement ecosystem.

annual reports, notices of investigation, implementation guides, technical advisories, and multimedia programs such as podcasts, among others.

While guidance may not seem as consequential as precedents, they help establish new thresholds for enforceable behavior. For example, in April 2020, the United Kingdom DPA released a report [22] that summarized how their office will regulate during the Coronavirus pandemic. In there, they emphasized that organizations should continue to meet the 72-hour deadline for reporting data breaches. The report also laid out a new priority: to take firm and swift action against those looking to exploit the public through nuisance calls or by misusing personal information in the guise of the pandemic. Similarly, the 2020 annual report of the UK DPA [5] indicated that out of the 1446 data breaches they investigated, 28.1% were because of *emailing or faxing personal data to incorrect recipients*. As is clear from these illustrative examples, even legal guidance helps determine the state of the art in GDPR enforcement.

3.3 Procuring the Information

GDPR, via $\mathscr{G}$57, $\mathscr{G}$59, and $\mathscr{G}$70 require DPAs and EDPB to make the aforementioned documents available to the public. Though the law does not mandate using the Internet as a platform for sharing such data, in practice, we have seen most of these agencies embrace the electronic format and posting content on their websites. This is critical for us since one of our goals is to operate the system without a human-in-the-loop. That said, we have encountered significant diversity in terms of website organization, document formats, languages employed, and frequency of updates across agencies, which have to be incorporated into our crawler.

3.4 Scope and Limitations of Our Approach

While this modeling of the GDPR ecosystem and the methodology to procure data does fulfill our project goals, it also results in some limitations in terms of scope and functionality. We address two such concerns here:

Why not include non-official sources and content? We acknowledge that SOTA can also be informed by non-official sources such as law journals, investigative news reports, cybersecurity research papers, technological breakthroughs, and white papers from companies, among others. For example, in 2020, Cohen and Nissim published results [16] demonstrating k-anonymization technique does not meet GDPR's requirement of not allowing singling-out on an anonymized personal data set. While such findings are potentially useful, expanding the scope beyond the official sources imposes two challenges: (i) the volume of data i.e., the number of secondary sources and the content they generate is significantly higher than those from the official agencies. For instance, just in the area of computer security and privacy, the total number of conferences and journals exceeds the number of DPAs by an order of magnitude [7]. (ii) The need to vet the information for accuracy and consistency. We are not aware of any automated means to determine the quality, relevance, and accuracy of information from such a broad range of sources. Thus, for the time being, we have decided to exclude these secondary sources. That said, we expect all significant findings to make their way into official GDPR enforcement documents albeit with a delay.

Would the resulting SoTA knowledgebase answer all of my GDPR questions? The goal of this project is simply to create and maintain a repository of GDPR enforcement knowledgebase.

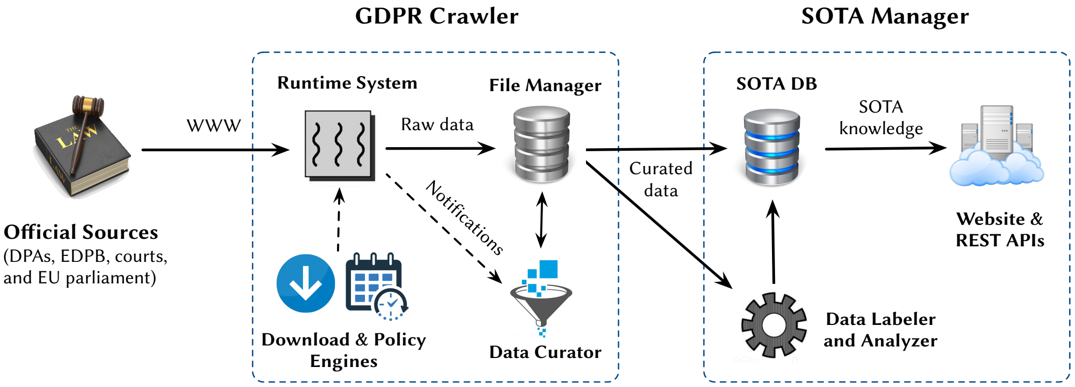


Figure 2: System architecture of GDPRxiv

As such, our system's knowledge is limited to only those aspects on which official GDPR bodies have deliberated up on or decided on. This approach results in two limitations. First, our system, GDPRxiv, would not be able to provide any information if the topic of interest does not appear in prior GDPR precedents or guidance. Continuing with the previous example on k-anonymization, our system would not tell if and when an organization should stop using k-anonymization, or suggest an alternative technique, or indicate if would result in a penalty. Second, even when the SOTA information exists for a user's question, our system does not provide any advice or additional insights; it simply provides access to the related SOTA-defining documents. The users of our system will have to draw their own conclusions. In contrast, there are other efforts that provide intelligent insights such as predicting the amount of GDPR fines [27] and automating GDPR compliance checking [12, 31].

4 GDPRxiv

In this section, we present the design and implementation of GDPRxiv; describe the technical challenges in crawling and sustaining this knowledge base; and outline the usability of the SOTA repository for a broad range of people in the computing community.

4.1 Crawler Design and Implementation

Informed by our GDPR enforcement model, and inspired by the WWW crawlers [9, 15, 23], we propose an architecture for the GDPR crawler as shown in Figure-2. It has five key components: (i) a *policy engine* that specifies crawl configurations like GDPR source list, crawling frequency, and status of crawled documents, (ii) a *download engine* that implements HTML parsing, URL extraction, and document downloading, (iii) a *data curator* that filters out non-GDPR documents, classifies files by type, and translates them to English, (iv) a *file manager* that administers the enforcement database including low-level access to files, and finally (v) a *runtime system* that manages cloud infrastructure, inter-component communications, and error handling.

We have implemented the whole system in Python and deployed it on Google cloud. The choice of Python was driven by its usability, large developer base, and extensive libraries, while that of Google cloud was due to its translation service and language processing capabilities. The download engine is built using `BeautifulSoup` and `Selenium` driver for paginating, identifying, and downloading files from the source websites. For every downloaded document, the policy engine keeps a reference including its <title, URL, release date, MD5>, so as to avoid duplicate downloads in the future. Next, the data curator employs `PyPDF` to convert the downloaded files into plain text format, then invokes Google `translate v2` APIs to generate English content. When the data curator finds a non-GDPR document, it informs the policy engine to add it to a `do-not-crawl` list. We have implemented all the crawler functionalities in 6076 lines of code and will open source the system after the peer review process.

Finally, since GDPR crawler does not experience the scale challenges of generic WWW crawlers (for example, our source list has ~30 websites[4] that generate an average of ~200 documents per month), we do not evaluate it against traditional performance metrics such as CPU and memory consumption, load on storage system, crawling throughput and latency, or the imposed load on the target websites; instead, we deliberate on the quality of the procured corpora.

4.2 Quality and Accuracy

Filtering non-GDPR documents. A number of DPAs existed and operated before GDPR, and continue to oversee other regulations in addition to GDPR. So, it is likely that some of the documents obtained by our crawler are non-GDPR ones. We employ a simple two-step filtering: first, we exclude all the documents dated prior to May 25th 2018; second, we omit documents that do not contain text such as GDPR, General Data Protection Regulation, EU 2016/679, one of the 99 articles by name, or a translated version of these phrases (for instance, Spain's version of GDPR is called *Reglamento General de Protección de Datos* or *RGPD* for short). The simplicity of our filtering heuristic could lead to false positives i.e., we end up adding non-GDPR documents that mention one of these words

[4] Our modeling in Section-3 requires us to crawl a variety of official sources. However, in practice, we find that the DPAs put out all the enforcement documents that involves them in any capacity. This includes laws passed by EU and the national parliaments, court cases involving DPAs, as well as EDPB decisions. Thus, it is sufficient to simply crawl DPA websites.

Table 3: ***Representative members of the CS community and their expected needs from the SOTA***

Computing community	Primary requirement (w/ example)
System designers	Avoiding non-compliant design choices *Does GDPR allow using CCTV feeds to determine student attendance?*
Programmers	Determining metrics and mechanisms compliant with the law *Does "Right to be Forgotten" require the data be purged from the backups?*
Systems administrators	Monitoring for issues in deployed systems *Notify me if Oracle DB gets mentioned in any enforcement*
Security researchers	Data-driven research and development *Are decentralized approaches more error-prone in managing consent?*
Consumers	Discovering the powers and limits of data rights *Could I ask my bank to not use AI for deciding on my loan application?*
Students	A platform for learning about cybersecurity *What are the real-world implications of "Right to be Forgotten"?*

in the passing. We have tested several random samples of ~25 documents to confirm that the false positive rate is never more than 5%. Our choice reflects a preference for safety (i.e., not missing a valid GDPR document) over accuracy (i.e., having a small number of non-GDPR documents).

Accuracy of translations. Most DPAs either do not have an English language website or do not link all the required GDPR documents in the English version of their website. So, our crawler procures documents in the native language and uses Google translate to convert them to English. This implies that the quality and accuracy of our repository is dependent on Google translate. This is a necessary limitation of our system. However, we note that Google Translate was initially trained using linguistic data from the European parliament (and the United Nations assembly).

4.3 Labeling the Enforcement Corpora

The enforcement documents are not only unstructured but also exhibit significant variance in their length, layout, and legal matter. So, we built a data labeling engine to identify the salient characteristics of each document. We store the labels as key-value pairs in a separate file but directly associated with the original document. First, the crawler itself generates a set of labels including the `country`, `origin language`, `issuing agency`, and `document type`. Next, we built an automated labeler using NLTK [10] that is able to extract information such as `word count`, `estimated read time`, `release date`, `quoted GDPR articles`, and `financial penalty`, if any. To ensure that our labeling routine works correctly, we performed these extractions manually on a random sample of ~25 documents (then fixed the issues and repeated this step iteratively until no more errors were left). Finally, the system also allows integrating human-generated or externally generated labels. For example, to each precedent document that had a financial penalty, we affixed a `business-size` label that marked whether the company is a micro, small, medium, or large business, or a government entity.

4.4 Disseminating the SOTA Knowledge

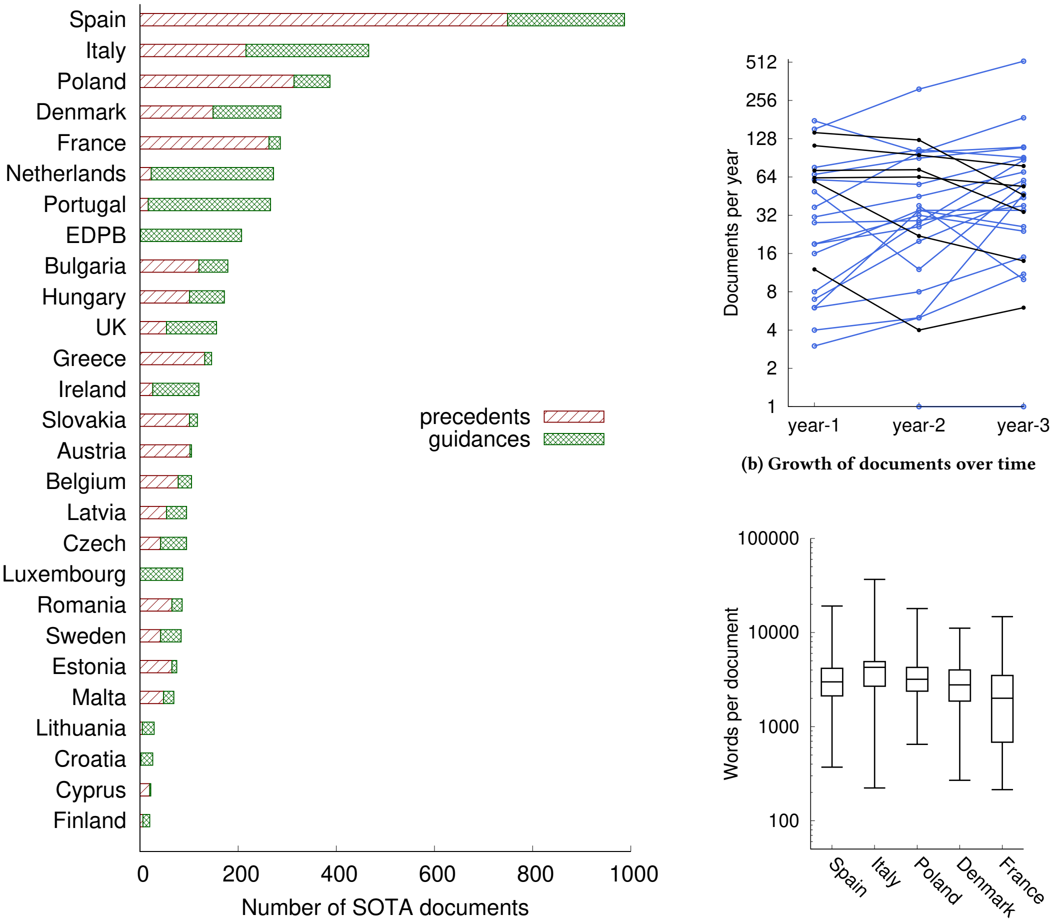
The last part of our system addresses accessibility and usability of the knowledge base. We intend GDPRxiv to be used as a first source of GDPR information by the computing community. Table-3 lists representative members of the community, and their anticipated use cases. While this is not exhaustive, it helps us build interfaces that would provide the broadest coverage. We discuss the evolution of GDPRxiv in more detail in Section-6.

GDPRxiv.org website. The website provides a search-based interface to the enforcement corpora followed by an option to filter the results by country, GDPR articles, penalty level, and other labels. Users can also access and bulk download the original documents. Lastly, the website provides insights and high-level summaries concerning the SOTA knowledge. We have built the search interface using React, a JavaScript library, and deployed it on Google AppEngine. The entire enforcement corpora and its associated label data are stored in a PostgreSQL database.

4.5 Sustainability

By definition, any knowledge considered SOTA will get stale if not continually updated and maintained. In our case, this translates to keeping the enforcement corpora up to date (i.e., operational continuity) and making sure that the crawler functionalities are not broken (i.e., project management continuity). Based on the data from first three years, we expect a crawling frequency of once per month to be a good compromise between keeping the SOTA fresh versus procuring a good volume of new data. We will configure Google Cloud Scheduler to run the crawler pipeline (including labeling) at this frequency.

However, it is more challenging to ensure that the crawler functionalities continue to work over time. This can happen because any of the DPAs could change their website layouts, modify the document formats, introduce restrictions on automated crawling,



(a) SOTA documents sorted by their source

(b) Growth of documents over time

(c) Spread in document size

Figure 3: Document-level characterization of SOTA: (a) distribution of documents across different countries, (b) growth in the volume of SOTA from first to third year, and (c) document sizes from the five most active DPAs.

or create new sources of information. Accommodating these, invariably leads to code changes in GDPRxiv, and thus human intervention. Many open-source research projects face this challenge and have found effective ways to sustain themselves beyond the initial research phase. For example, Spark framework [1] developed at the University of California Berkeley was donated to Apache Software Foundation; DAWNbench [17] created at Stanford University formed a consortium with stakeholders from industry; and GDPRhub [29] has directly engaged with the members of the community. We plan to explore this in the future.

5 INSIGHTS AND TRENDS

Having access to the entire[5] GDPR corpora enables us to make aggregate and longitudinal characterization of enforcement trends as well as perform fine grained content level analysis. We share key insights and trends from the first three years of GDPR roll out.

5.1 Document-level Trends

SOTA distribution. Figure-3a shows the official sources of GDPR information along the Y-axis and the aggregate number of SOTA documents they have produced along the X-axis. The plot distinguishes the corpora into two bins: legal precedents (in red) and guidances (in green). Starting from the top, we see a steep drop in the way GDPR enforced and interpreted for the stakeholders. In fact, the top five DPAs account for 49% of the documents while the bottom half contribute only 20%. Next, we also see divergence in the way DPAs operate: some are more proactive (i.e., they generate more guidances for stakeholders compared to issuing decisions and penalties) whereas others are reactive (i.e., they wait until complaints are received and then take action). Overall, we observe a 56:44 split between precedents and guidances.

SOTA over time. Figure-3b shows how SOTA documents are generated over time. The X-axis indicates the three years of GDPR (i.e., May-2018 to May-2021), and the Y-axis measures the number of

[5] We must note that our dataset excludes two countries: Slovenia (whose parliament has not officially adapted GDPR yet) and Germany (due to a crawler interface issue with their DPAs). We are engaging with the DPA to resolve this.

Fine (in €)	Data controller	DPA	Date
50,000,000	Google	France	Jan 21, 2019
27,800,000	Gruppo TIM	Italy	Jan 15, 2020
22,046,000	British Airways	UK	Oct 16, 2020
20,450,000	Marriott	UK	Oct 30, 2020
16,700,000	Wind Tre	Italy	Jul 13, 2020
12,251,601	Vodafone	Italy	Nov 12, 2020
8,500,000	Eni Gas e Luce	Italy	Dec 11, 2019
8,150,000	Vodafone	Spain	Mar 11, 2021
6,000,000	Caixabank	Spain	Jan 13, 2021
5,000,000	BBVA	Spain	Dec 11, 2020
5,000,000	Google	Sweden	Mar 11, 2020

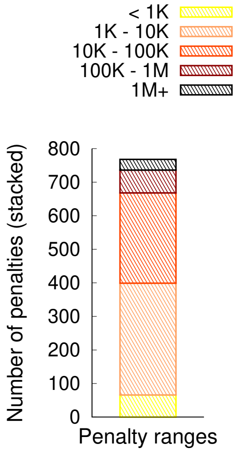


Figure 4: Characterizing financial penalties: (a) table lists the top-10 highest individual penalties, and (b) graph shows the distribution of penalties by penalty ranges (measured in €)

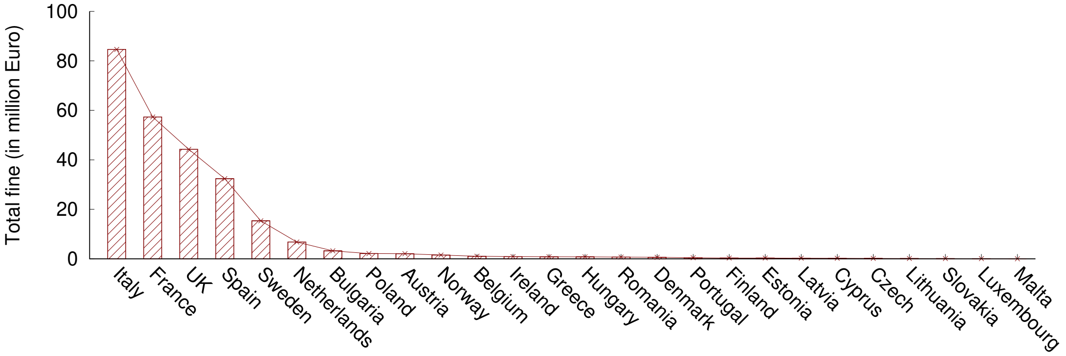


Figure 5: Distribution of financial penalties levied by EU nations

documents produced in a given year. Each line in the graph corresponds to a DPA, with those DPAs whose year-3 count exceeds the year-1 count being plotted in blue, and the rest being plotted in black. First off, we see that more than two-thirds of the DPAs have expanded on their SOTA generation, indicating increased enforcement and outreach activities. In raw numbers, year-3 corpus was 52% more than the year-1 corpus.

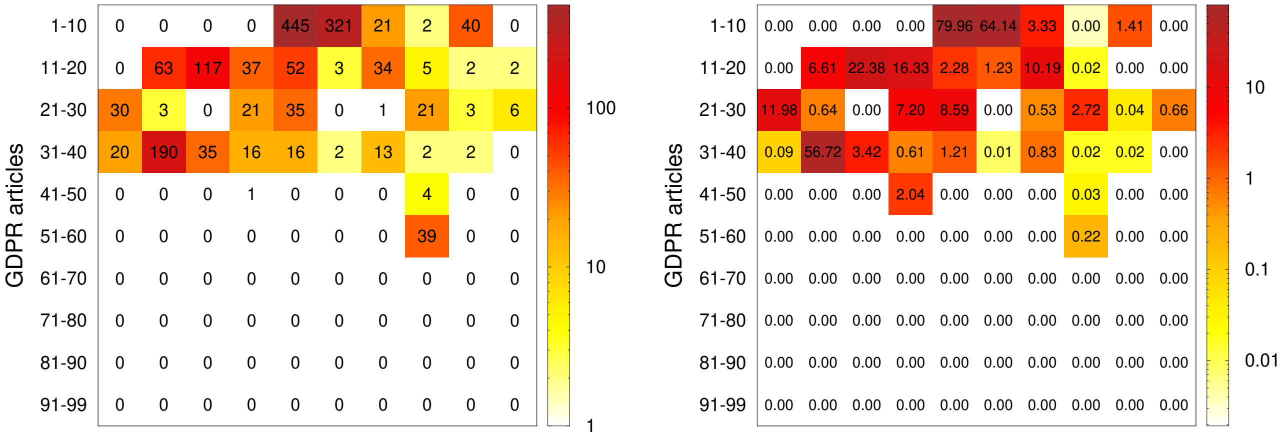
SOTA content. Finally, figure-3c characterizes the SOTA documents in terms of their content length. Using candlestick plot, Y-axis represents the word count of all the documents that belong to a given DPA, which is indicated along the X-axis. The base of the candlestick box represents the 25th percentile, its top represents the 75th percentile while its midpoint denotes the mean word count. The minimum and maximum word counts are marked by the whiskers on either side. While a majority of the documents range between 500 to 5000 words (i.e., the area inside the boxes), we see significant variation in the other half of the corpora. Such a divergence is not unexpected since these documents vary in their language, layout, legal matter, target audience, time of publication, among other things.

In summary, this high-level characterization of the corpora reveals three novel insights:

- *Not all DPAs enforce GDPR equally.* We show that just five DPAs account for 49% of all GDPR activity; and we identify how certain DPAs are proactive in issuing guidances to stakeholders versus others who reactively issue penalties.
- *GDPR activities are increasing over the years.* Our analysis shows that the number of SOTA-defining documents increased by 52% from the first to third year of GDPR.
- *GDPR SOTA is diverse.* Our SOTA collection indicates a 56:44 split between precedent and guidance documents, and that these documents vary considerably in their length, language, layout, and legal matter.

5.2 Content-based Insights

To get a deeper understanding of GDPR's enforcement, we analyze the content of the corpora. Specifically, we investigate the (i) impact of enforcement (by studying the financial penalties), (ii) areas of



1-10	0	0	0	0	445	321	21	2	40	0
11-20	0	63	117	37	52	3	34	5	2	2
21-30	30	3	0	21	35	0	1	21	3	6
31-40	20	190	35	16	16	2	13	2	2	0
41-50	0	0	0	1	0	0	0	4	0	0
51-60	0	0	0	0	0	0	0	39	0	0
61-70	0	0	0	0	0	0	0	0	0	0
71-80	0	0	0	0	0	0	0	0	0	0
81-90	0	0	0	0	0	0	0	0	0	0
91-99	0	0	0	0	0	0	0	0	0	0

(a) Number of times each article is cited

1-10	0.00	0.00	0.00	0.00	79.96	64.14	3.33	0.00	1.41	0.00
11-20	0.00	6.61	22.38	16.33	2.28	1.23	10.19	0.02	0.00	0.00
21-30	11.98	0.64	0.00	7.20	8.59	0.00	0.53	2.72	0.04	0.66
31-40	0.09	56.72	3.42	0.61	1.21	0.01	0.83	0.02	0.02	0.00
41-50	0.00	0.00	0.00	2.04	0.00	0.00	0.00	0.03	0.00	0.00
51-60	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.22	0.00	0.00
61-70	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
71-80	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
81-90	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
91-99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

(b) Fines attributed to each article (in million €)

Figure 6: Heatmaps showing (a) frequency of citation and (b) penalties for each article of GDPR

enforcement (by tabulating the cited GDPR articles), and (iii) targets of enforcement (by examining the data controllers). For these analyses, we limit our dataset to only those cases that involve a financial penalty.

Financial Penalties. GDPR grants DPAs broad authority in levying financial penalties on data controllers that fail to comply with GDPR. While 𝒢83 lays out a detailed set of conditions for assessing the severity of the infringement, it leaves it up to the DPA to determine the amount of resulting penalty (by only setting the maximum limit to €20M or 4% of the organization's worldwide annual turnover, whichever is higher). The goal of this analysis is to shed light on how penalties have manifested in the field.

There were a total of 768 penalties in the first three years of GDPR (ending on May 25, 2021), which translates to a penalty once every 1.42 days on average. Figure-5 shows the distribution of financial penalties with X-axis marking the DPAs while the Y-axis representing the aggregate fines over the three years. Similar to the GDPR activities graph (Figure-3a), we see that a handful DPAs are heavy finers while the majority have not been that punitive. In fact, the top five DPAs are responsible for 76% of all fines issued while the bottom half only accounts for 0.01%.

Next, table in Figure-4 lists the top 10[6] penalties across all DPAs. Expectedly, we find that all these have been issued by one of the top five DPAs from Figure–5. Another trend, which is also seen in the DPA activities graph (Figure-3b), is that enforcements have picked up pace in year three compared to the first two years. For instance, seven fines in the top-10 table were issued in year three of GDPR (i.e., between May 2020 and May 2021).

Finally, graph in Figure-4 depicts the distribution of penalties by their penalty ranges. We group all 768 penalties in five ranges starting with <1K Euros and increasing the amount by one order of magnitude until we reach >1M Euros. We see that 52% of penalties were for less than €10K, and only 13% were for more than €100K.

Cited GDPR Articles. The articles of GDPR, which are 99 in number, could be grouped into five broad categories. 𝒢1-11 contain definitions and principles of personal data processing; 𝒢12-23 establish the rights of the people; 𝒢24-50 mandate the responsibilities of the data controllers and processors; the following 26 articles describe the role and tasks of the data protection authorities; and the remainder of them cover liabilities, penalties and other specific situations. DPAs could rely on any number of these articles to carry out their enforcements. The goal of this analysis is to identify the areas of focus by tracking the articles cited in enforcements.

Figure-6a shows a heatmap that represent each of the 99 articles as boxes and with each box being colored in proportion to the number of times the corresponding article is cited. As shown in the adjacent heatmap scale, the lighter hue of yellow indicates low citations whereas the darker hues of orange and then red indicate higher citations. The clear bifurcation between the first and second halves of the articles is no surprise since the first 50 articles cover the core data management principles, the rights of the people and the expected behavior of organizations – the kinds of articles that could form the basis for establishing GDPR violations. However, the spread of citations within the top half of the articles offers two interesting takeaways: there is a heavy skew with three articles 5, 6, and 32 such that at least one of them appears in 78.5% of all citations; contrary to the popular media coverage, reporting of data breaches (𝒢33-34) or prominent rights such as *Right To be Forgotten* (𝒢17), *Right to Object* (𝒢21), and *Right of Access* (𝒢15) have not resulted in a significant number of citations.

To further validate these observations, we analyze how infringement of specific articles results in different financial penalties. To arrive at this distribution, we iterate through all the enforcements, dividing the imposed financial penalty equally amongst all the cited articles within that enforcement. The resulting heatmap is plotted

[6] the list has 11 entries due to a tie for the 10th largest fine

Table 4: ***Key articles of GDPR that represent the focus areas of enforcements***

No	GDPR article and its key clauses	What they regulate (paraphrased)
5	Principles relating to processing of data	
	(1b) Purpose limitation	Collect data for explicit purposes
	(1c) Data minimization	Collect only minimally necessary data
	(1e) Storage limitation	Do not store data indefinitely
	(2) Accountability	Be able to demonstrate compliance
6	Lawfulness of processing	
	(1) Conditions to establish lawfulness	Six conditions including obtaining consent from the data subject; establishing the necessity of data collection and processing; and so on
	(2) Data usage beyond initial purpose	Four conditions including establishing a link between the two purposes; analyzing the consequences of new purpose;
32	Security of processing	
	(1) State of the art	Implement security measures that match the state of the art in the field
	(2) Proportionality	Implement security measures in proportion to type of data being processed

in Figure-6b, where each of the 99 boxes represent an article and the color of each box measures the aggregate fine attributed to that article. The heatmap scale on the right shows lower fines in lighter yellow hue and higher fines in darker orange/red hues (fines are measured in million €). We see that the high level takeaways from Figure-6a hold good with the same three articles (5, 6, and 32) accounting for 65% of all financial penalties.

Given the importance of articles 5, 6, and 32, we would be remiss not to have a discussion on them. Table-4 presents an accessible description of these articles by highlighting their key clauses and by explaining how these translate to the computing domain. The focus on these articles conveys the importance that regulator are placing on sound data management practices starting from how personal data is to be procured (𝒢5), how it is to be processed (𝒢6), and how the security infrastructure is to be designed and operated (𝒢32). A deeper analysis of the top-10 highest penalties reveals this approach of DPAs: it is not the actual data breaches or the unintentional violations of people's rights that gets huge penalties, but rather a lack of responsible data management systems and processes underneath.

Data Controllers. By law, GDPR applies to all entities that collect and process personal data irrespective of their size and capacity i.e., an individual offering photography services is bound by the same regulation as that of say, Google and Facebook. However, GDPR does limit the maximum penalty that can be levied on a data controller to be 4% of their annual revenue or €20M (via 𝒢83). This setting gives DPAs a broad autonomy in determining which violations to investigate and then how much penalty to impose. The goal of this analysis is to understand the impact of enforcement across different categories and types of companies.

We iterate through all the financial penalties imposed in the first three years, and classify the targeted data controllers into micro, small, medium, and large companies (we separately tabulate government agencies and undisclosed recipients). To arrive at this classification, we use the SME definition of the EU Commission [18], and then look up the company profile in Dun and Bradstreet business directory [21]. The resulting data is summarized in Table-5, where columns three and four list the total number and amount of fines respectively.

Table 5: ***Tabulating penalties by data controller categories***

Category	Max staff size, max revenue	Number of fines	Amount of fines
Micro	10 and €2M	166	€1.14M
Small	50 and €10M	93	€1.44M
Medium	250 and €50M	49	€3.84M
Large	—	260	€286.71M
Government	—	115	€9.84M
Undisclosed	—	85	€2.47M

We see that large companies have received 260 fines (or 34% of all fines), which is roughly the same as small and micro companies put together. More generally, the SME category (i.e., small, medium, and micro companies), despite accounting for 99% of all businesses in the EU, received only 40% of all penalties. Since DPAs do not reveal any details about their investigation policies nor on the complaints that did not result in penalties, it is hard to determine the reason for this skew. However, what is clear from the fourth column is that DPAs have put in practice GDPR's provision on *proportional penalty*. Large companies have paid 94% of the total amount of fines despite receiving only a third of the penalties. While the average penalty on an SME is €20.8K, the large ones pay €1.1M on average. This is a clear sign that it is not only the infringement that counts but also the scale and scope of its impact (presumably, large companies hold significantly high amount of personal data, thereby affecting larger population).

6 VISION AND OPPORTUNITIES

Comprehending and complying with *emerging data rights regulations* is a challenging socio-technical problem. By creating a reliable source of ground truth in the form of GDPR SOTA, our work opens up a broad range of opportunities.

A platform for privacy enhancing applications. One of our goals is to evolve this system into a data platform upon which user-specific applications could be built. For instance, think of a *web browser plugin* that warns you of GDPR compliance issues as you visit websites; or a *notification service* that sends out an email when a new GDPR enforcement document that matches a specified criteria (say CCTV, Oracle DB, or article-13) gets posted in the SOTA repository. Towards this goal, we have defined a set of REST APIs that allow programmatic access to GDPRxiv. We are currently implementing a prototype RESTful service in Python using the Flask framework. We will include detailed documentation about GDPRxiv REST APIs in the project website.

An educational tool for data rights. We envision GDPRxiv being adapted as a pedagogical tool for data-driven education and exploration of GDPR and similar regulations. Specifically, GDPRxiv could enable students (i) to acquire a working knowledge of GDPR in the wild, (ii) to understand how GDPR impacts the design, development, and deployment of computing systems, and (iii) to build tools and services on top GDPRxiv's programming interface. We draw inspiration from prior work such as Azure VM Traces [20], Google cluster traces [26], and Million privacy policies [8] that have been widely used in academic settings and research environments.

Predicting and preparing for compliance challenges. GDPRxiv data indicates that in the first three years of GDPR, regulators have levied penalties on 600+ organizations amounting to €305M. Interestingly, recent estimates put the initial compliance cost spent by companies operating in EU to be in the range of few billion euros. At this point, it is apparent that achieving GDPR compliance is not a one-time effort but a constant journey towards improving the security and privacy of personal data systems. Thus, it is important to follow and project the enforcement trends, and then prepare for upcoming compliance tasks. We envision GDPRxiv to be a platform on which techniques from natural language processing (NLP), machine learning (ML), and computer human interaction (CHI) could be deployed to obtain actionable and more advanced insights.

Bridging the gap between law and CS. We are in the early and formative days of personal-data rights. There is an implicit tussle between the legal and computing communities on how to define, implement, and enforce these rights. When legal scholars draft regulations, they tend to focus on the core legal principles that are broadly interpretable and that would hold the test of time, instead of getting into the specific implementations of the current time. While legally prudent, this approach is in stark contrast with the established practices of the computing community that expects precise specifications to build and operate computing systems. We believe that GDPRxiv has the potential to reduce the tussle between the two communities by creating a *lingua franca* for exchanging information and offering feedback.

7 CONCLUSION

In this work, we make an argument that having a well understood SOTA is paramount for an evolving field like data rights, of which GDPR is the most visible legal-computing standard. We define what the SOTA for GDPR should be, and propose a methodology to compose it. Then, we design and implement an information archiver system called *GDPRxiv* that collects, curates, organizes, disseminates, and sustains the SOTA-defining documents. We have put together the largest centralized collection of GDPR knowledge base, and we envision GDPRxiv to evolve into a platform for data-driven education and research concerning GDPR compliance and enforcement.

REFERENCES

[1] 2014. The Apache Software Foundation Announces Apache Spark as a Top-Level Project. https://blogs.apache.org/foundation/entry/the_apache_software_foundation_announces50.

[2] 2018. California Consumer Privacy Act. *California Civil Code, Section 1798.100* (Jun 28 2018).

[3] 2018. Data Deletion on Google Cloud Platform. https://cloud.google.com/security/deletion/.

[4] 2018. General Law for the Protection of Personal Data (LGPD). *Brazil statutory law 13.709* (Aug 14 2018).

[5] 2020. Data Security Trends 2020-21 Q1. https://ico.org.uk/action-weve-taken/data-security-incident-trends/.

[6] 2021. Consumer Data Protection Act. *Virginia Acts of Assembly, 2021 Special Session I* (Mar 2 2021).

[7] 2021. Cybersecurity Conferences 2021 - 2022. https://infosec-conferences.com/.

[8] Ryan Amos, Gunes Acar, Elena Lucherini, Mihir Kshirsagar, Arvind Narayanan, and Jonathan Mayer. 2021. Privacy Policies over Time: Curation and Analysis of a Million-Document Dataset. In *ACM WWW*.

[9] Arvind Arasu, Junghoo Cho, Hector Garcia-Molina, Andreas Paepcke, and Sriram Raghavan. 2001. Searching the Web. *ACM Transactions on Internet Technology (TOIT)* 1, 1 (2001), 2–43.

[10] Steven Bird and Edward Loper. 2004. NLTK: The Natural Language Toolkit. In *Proceedings of the ACL Interactive Poster and Demonstration Sessions*. https://www.aclweb.org/anthology/P04-3031

[11] European Data Protection Board. 2021. Our Documents. https://edpb.europa.eu/about-edpb/about-edpb/who-we-are_en.

[12] Piero A Bonatti, Sabrina Kirrane, Iliana M Petrova, and Luigi Sauro. 2020. Machine Understandable Policies and GDPR Compliance Checking. *KI-Künstliche Intelligenz* 34, 3 (2020), 303–315.

[13] Anu Bradford. 2020. *The Brussels effect: How the European Union rules the world*. Oxford University Press, USA.

[14] Julie Brill. 2018. Microsoft's commitment to GDPR, privacy and putting customers in control of their own data. In *Microsoft Blog*.

[15] Sergey Brin and Lawrence Page. 1998. The Anatomy of a Large-Scale Hypertextual Web Search Engine. In *ACM WWW*.

[16] Aloni Cohen and Kobbi Nissim. 2020. Towards formalizing the GDPR's notion of singling out. *PNAS* 117, 15 (2020), 8344–8352.

[17] Cody Coleman, Deepak Narayanan, Daniel Kang, Tian Zhao, Jian Zhang, Luigi Nardi, Peter Bailis, Kunle Olukotun, Chris Re, and Matei Zaharia. 2017. DAWNBench: An End-to-End Deep Learning Benchmark and Competition. In *NIPS ML Systems Workshop*.

[18] European Commission. 2021. SME definition. https://ec.europa.eu/growth/smes/sme-definition_en.

[19] Brian Cooper, Adam Silberstein, Erwin Tam, Raghu Ramakrishnan, and Russell Sears. 2010. Benchmarking Cloud Serving Systems with YCSB. In *ACM SoCC*.

[20] Eli Cortez, Anand Bonde, Alexandre Muzio, Mark Russinovich, Marcus Fontoura, and Ricardo Bianchini. 2017. Resource central: Understanding and predicting workloads for improved resource management in large cloud platforms. In *ACM SOSP*.

[21] Dun and Bradstreet. 2021. Business Directory. https://www.dnb.com/business-directory.html.

[22] Information Commissioner Elizabeth Denham. 2020. How We Will Regulate During Coronavirus. In *UK ICO Blog*. https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2020/04/how-we-will-regulate-during-coronavirus/.

[23] Jonathan M Hsieh, Steven D Gribble, and Henry M Levy. 2010. The Architecture and Implementation of an Extensible Web Crawler. In *USENIX NSDI*.

[24] CMS Law. 2021. GDPR Enforcement Tracker. https://www.enforcementtracker.com/.

[25] General Data Protection Regulation. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46. *Official Journal of the European Union* 59, 1-88 (2016).

[26] Charles Reiss, John Wilkes, and Joseph L Hellerstein. 2011. Google cluster-usage traces: format+ schema. *Google White Paper* (2011).

[27] Jukka Ruohonen and Kalle Hjerppe. 2020. Predicting the Amount of GDPR Fines. *arXiv preprint arXiv:2003.05151* (2020).

[28] Adam Satariano. 2019. Google is fined $57 Million Under Europe's Data Privacy Law. In *The New York Times*. https://www.nytimes.com/2019/01/21/technology/google-europe-gdpr-fine.html.

[29] Max Schrems. 2021. GDPRhub. https://gdprhub.eu/.

[30] Supreeth Shastri, Vinay Banakar, Melissa Wasserman, Arun Kumar, and Vijay Chidambaram. 2020. Understanding and Benchmarking the Impact of GDPR on Database Systems. *PVLDB* 13, 7 (2020), 1064–1077.

[31] Damiano Torre, Ghanem Soltana, Mehrdad Sabetzadeh, Lionel C Briand, Yuri Auffinger, and Peter Goes. 2019. Using models to enable compliance checking against the GDPR: an experience report. In *2019 ACM/IEEE MODELS*.