GDPRxiv: Establishing the State of the Art in GDPR Enforcement

Chen Sun
Computer Science
University of Iowa

Evan Jacobs
Computer Science
University of Iowa

Daniel Lehmann
Computer Science
University of Copenhagen

Andrew Crouse
College 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 8000+ official precedents and guidances, the largest such collection. To demonstrate the usefulness of this corpora, we share insights gleaned at the aggregate-level (say, how is the GDPR being enforced in the field) and at the article-level (say, what are the common failures observed in the field while implementing article-17 Right to be Forgotten). We release all of our software artifacts and datasets at https://GDPRxiv.org.

KEYWORDS

GDPR, GDPR enforcement, GDPR SOTA

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) [2] 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 [3, 5, 7, 18]. 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, policies, 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 SOTA is at any given time. Thus, the goal of our work is to build such a knowledge base for 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. Consider the journey our society has gone through for women's rights or civil rights; GDPR and personal data rights are just four 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, and financial penalties. Table-1 provides a concise summary of these findings. We also demonstrate how to analyze GDPRxiv corpora from a computing perspective, resulting in the identification of a dozen plus common failures for article-17 the *Right-to-be-Forgotten*. While four years is a short time to judge the efficacies of a transformative regulation like GDPR, our findings do reflect that its enforcement is broadly aligned with its original intent. Our long-term vision is to evolve GDPRxiv into a platform for data-driven research and analysis of 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:

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

Proceedings on Privacy Enhancing Technologies 2023(4), 484–499

https://doi.org/10.56553/popets-2023-0121

Enforcement activities	GDPR is not implemented uniformly across Europe. *Three countries (ESP, DNK, POL) account for 54% 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 enforcement decisions than year-1.*
Financial penalties	GDPR's proportional penalty has resulted in a heavy skew in the application of fines. *80% of the fines were for <€10K and only 1.8% violators ended up with million € fines.* *Three countries (LUX, IRL, FRA) are responsible for 83% of all GDPR fines.*
Focus areas	Regulators are prioritizing sound & 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 nearly 50% of all citations.*

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

- **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 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 5×. We publicly release all our software artifacts and datasets at https://GDPRxiv.org.
- **Tracking GDPR in the Wild:** We provide a longitudinal characterization of GDPR's enforcement in the wild, identifying the presence of proportional penalties, and characterizing the skew in its application across Europe.
- **RTBF Analysis:** We analyze all RTBF-related enforcements from a computing perspective to chronicle a dozen plus RTBF failures commonly seen in the field. Our Via Negativa approach demonstrates the power of GDPRxiv in identifying failure patterns, and thereby providing actionable insights to the computing community.

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 [2] 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) [3] and Virginia's Consumer Data Protection Act (CDPA) [7] 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 [18] that it would voluntarily extend the core rights of GDPR across the world. Legal scholars refer to this phenomenon as *the Brussels effect* [17], a race to the top effect where the early but stringent standards of an EU regulation get 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 women's rights or 1960s were for 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 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

[2]henceforth, we will prefix GDPR articles with 𝒢

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 [47] Google €50M for lacking legal basis for such purpose bundling. If we take 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 [50] 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 $\mathcal{G}$ 17 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 [50] 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 thorough deletion of customer data from all their systems but requires up to 180 days to complete the operation [4]. There are many other design parameters to deletion and other GDPR requirements that amplify uncertainty at design level.

Example-3: Uncertainty at operations level. *GDPR, via $\mathcal{G}$ 30 Records of Processing Activities and $\mathcal{G}$ 33 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 [50] has shown that for realistic workloads such as YCSB [24], 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 real-time 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) the complexity of GDPR enforcement 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 (roughly, 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 must 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. 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** [37]: is a website and mobile app that displays penalties levied under GDPR. As of Oct-2022, it consists of 1430 entries covering data protection authorities from all EU nations. Its key shortcomings vis-a-vis our effort are: (i) keeping its data collection and analysis methods

[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.

	Enforcement Tracker	GDPRhub	GDPRxiv
Collection method	Proprietary	Hand-curated by volunteers	Open-source crawler
Content types	Enforcements with penalties	All enforcements; Court judgements	All enforcements; Court judgements; Official opinions, reports, and guidance
No. of documents	1428	1594	8943
Interfaces	Website; Mobile app	Wiki; Newsletter	Website; REST APIs (in progress)
Sustainability	Unknown	Needs person-hours proportional to the documents added	Fully automated

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

proprietary, and (ii) focusing only on cases where monetary penalties are involved.

- **GDPRhub** [48]: is a wiki-style information portal, populated by voluntary contributors, that provides commentaries on GDPR enforcement. As of Oct-2022, it describes ~1000 decisions from courts and data protection agencies. The main issue with GDPRhub is that, like Wikipedia, it cannot be considered a reliable source of ground truth since the quality and quantity of its content are governed by the availability and skill level 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 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, policies, configurations, and operational practices that fail the current legal standards of GDPR compliance. Our definition of GDPR SOTA is an example of the *via negativa* approach: instead of providing a recipe for how to comply with GDPR, it lays out all the different ways in which organizations have failed to comply with GDPR. The rest of this section describes our approach to composing this SOTA by identifying all the official sources that generate enforcement information and then proposing a way to procure them 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 [15]. 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

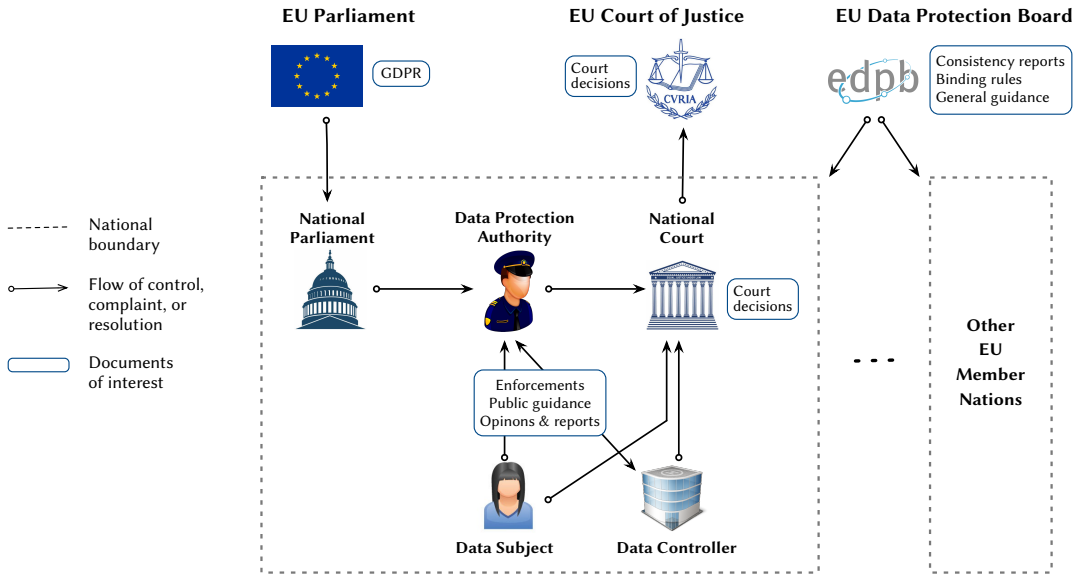


Figure 1: A representation of the GDPR enforcement ecosystem.

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; while examples of guidance include EDPB legal guidance, DPA's 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 precedent, they help establish new thresholds for enforceable behavior. For example, in April 2020, the UK DPA released a report [27] that summarized how their office will regulate during the Coronavirus pandemic. In there, they emphasized that during the pandemic, organizations should continue to meet the 72-hour deadline for reporting data breaches. They 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, their 2020 annual report [6] 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 examples, even legal guidances help 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 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

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 [21] 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 [8]. (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 enforcement documents albeit with a delay.

Would this SOTA answer all of my GDPR questions? The goal of this project is simply to create and maintain a repository of GDPR enforcement knowledge base. As such, our system's knowledge is limited to 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

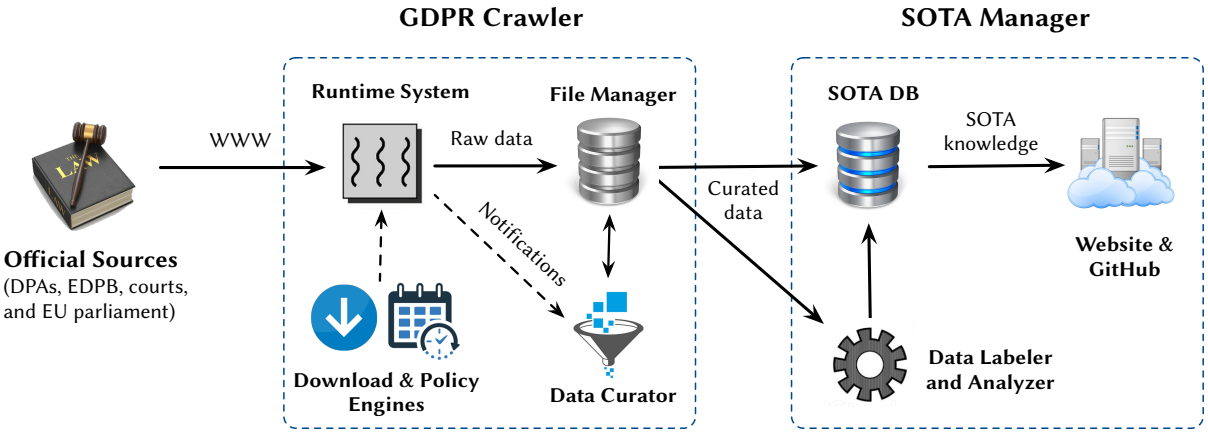


Figure 2: System architecture of GDPRxiv

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 [45] and automating GDPR compliance checking [16, 53].

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 its usability 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 [11, 19, 33], 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 our 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 document, the policy engine keeps a reference including its `<title, URL, release date, hash>` 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` 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 ~15K lines of code and will open source the system after the peer review process.

Performance. The GDPR crawler does not experience the scale challenges of generic WWW crawlers (for example, our source list has only 47 websites[4]). As of Nov-2022, the crawler takes about 22 hours to build the full repository from scratch when running on a Google Cloud n2-standard-4 VM (4 vCPU, 16GB RAM, 1TB SSD, 10Gbps network). This is due to our crawling being sequential, and our intention to minimize the workload on the DPA web servers by introducing a wait time of two seconds after every document download. A subsequent run, where no new documents are procured, completes within an hour. We posit that a crawling frequency of once per month should suffice (since, on average, it takes a month for ~200 documents to be generated). Given this use case, we did not undertake any further performance optimizations.

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 keywords such as GDPR, General Data Protection Regulation, EU 2016/679, one of the 99 articles by name, or a translated version

[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 the DPA websites.

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 in the passing. We have tested a random sample of 25 documents for every DPA 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 do not have English language websites, and even if they have one, they do not link all the required GDPR documents in the English language. So, our crawler procures documents in the native language and uses Google Translate to convert them to English. Our choice of Google Translate is motivated by (i) its generality and language coverage, and (ii) the fact it was initially trained using documents from the European parliament and the United Nations assembly [52]. However, this implies that the quality and accuracy of our repository is dependent on Google Translate. While Google Translate's accuracy has continued to improve [34, 55], several independent studies have highlighted its shortcomings [35, 54]. This is a current limitation of our system. However, if a better translation engine were to emerge, it would be straight forward to replace the current translation APIs in GDPRxiv (as well as rerun the previous translations).

4.3 Labeling the Enforcement Corpora

We built a labeling engine that identifies several key characteristics of the SOTA documents including the `country`, `origin language`, `issuing agency`, `document type`, and `release date`. We store these as key-value pairs in a JSON file associated with the original document. In addition, we have manually labeled four enforcement-specific metadata namely, `decision type`, `targeted organization`, `cited GDPR articles`, and `financial penalty` for all the precedent documents. This process of manual labeling was necessary since our efforts to automatically recognize these metadata did not produce accurate results. In particular, we tried (i) hand coded rules based on regular expressions, (ii) NLTK's built-in Named Entity Recognition [14], and (iii) BERT, a language model based on transformers [26]. We found that all these approaches produce low precision and recall, typically in 50-70% range. Since our goal is to accurately catalog the GDPR SOTA (and not to provide any statistical prediction/analysis), we chose the accuracy of (manual) labeling over ease of (automated) labeling.

While our choice mirrors that of other enforcement archives including GDPRhub and GDPR EnforcementTracker, we do acknowledge that this is a barrier for fully automating our workflow. That said, classification, extraction, and labeling of privacy law and policy documents is an emerging area of research [31, 32], and some researchers [38, 40] have recently demonstrated how the accuracy of NLP models could be improved with domain specific training. In particular, a team from the University of Sheffield and Athens University have trained LEGAL-BERT [20], which improves on the original BERT performance by pre-training it with legal corpora. We leave automated labeling as a future work.

4.4 Usability and Sustainability

Disseminating the SOTA Knowledge. We intend GDPRxiv to be used as a first source of GDPR information by the computing community. Our public website will provide 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 will also be able to access and bulk download the original documents. Lastly, the website would provide insights and high-level summaries concerning the SOTA knowledge.

Automating the GDPRxiv Processing Pipeline. By definition, any knowledge considered SOTA will get stale if not continually updated. In contrast to prior work [37, 48], GDPRxiv does offer the ability to automate this process completely. Our processing pipeline has three main tasks: (1) crawling the reference sites to identify new documents and then downloading them, (2) labeling the new corpora, and (3) reflecting the updates in the website. We are glad to report that all three steps are push-button automated i.e., a single CLI will execute each of the tasks, transparently and fully. However, it must be noted that step 2 only produces labels that can be automatically generated, and it will omit categories of metadata that are manually labeled. This limitation stems from our design choice of valuing the accuracy of labels but unable to find a technique that achieve ~100% accuracy. It is our intention to continue labeling the missing metadata manually, while exploring alternative approaches towards automating this step fully. For example, we are considering engaging with the EU regulatory bodies such as the EU Data Protection Board and the national DPAs to explore the feasibility of them providing a metadata file for every GDPR SOTA document they create. This would not only distribute the labeling burden but also result in labeled data of highest quality.

Long-Term Sustainability. Another challenge for GDPRxiv is the long term sustainability of the codebase i.e., ensuring that the crawler works despite any changes in the target websites. To put this into perspective, consider that in the first 4.5 years of GDPR, two DPAs have redesigned their websites/pages. So, we have not found this to be taxing. That said, many open-source research projects face this challenge and have found different ways to sustain themselves beyond the initial research phase. For example, the Spark framework [1] developed at Berkeley was donated to Apache Software Foundation; DAWNbench [23] created at Stanford formed a consortium with stakeholders from the industry; and GDPRhub [48] has directly engaged with the members of the community. We plan to explore these paths with inputs from the community.

Future Work. Finally, we identify the aspects of our system that would benefit from more research and development. First, *evaluating the user interfaces*. This is especially important since our goal is to make GDPRxiv useful for a broad set of stakeholders. Second, *automating the labeling engine*. This would not only reduce the maintenance burden but also enhance reproducibility. Third, *understanding the translation quality*. We have not yet done any quality check on Google Translate's output. Finally, expanding the source list. It would be worthwhile exploring the feasibility of including non-official sources of GDPR information, and how it affects the composition of SOTA.

5 EVALUATION

To evaluate the relevance and usefulness of our project, we ask and answer the following questions:

- How well does GDPRxiv compare against other GDPR corpora? (in Section-5.1)
- Does GDPRxiv help identify key trends in GDPR enforcement? (in Section-5.2)
- Can GDPRxiv help reduce uncertainties that computing community faces in complying with GDPR? (in Section-5.3)

Dataset. Having access to the entire official GDPR corpora allows us to perform aggregate characterizations as well as fine-grained analysis. The dataset for our evaluation includes all documents dated between 25-May-2018 and 24-Nov-2022 (i.e., *first four and half years* of GDPR). When comparing against GDPR Enforcement Tracker and GDPRhub, we use the same time period as well.

5.1 Accuracy and Scope

For this evaluation, we consider two measures: accuracy and scope. We consider a GDPR SOTA repository to be *accurate* if it is free of false content. Establishing the accuracy of GDPRxiv is easy: it follows directly from our data collection methodology (as discussed in Section-3). By restricting the sources of our data to official regulatory bodies, GDPRxiv is guaranteed to not contain any documents that are misinforming, disinforming, or malicious. The same cannot be said about the prior work: GDPRhub (which is a crowd-sourced wiki portal) or GDPR Enforcement Tracker (whose sources include both official and non-official publications).

Next, *scope* indicates the type(s) of SOTA defining documents contained within the GDPR SOTA repository. To characterize this, we compare GDPRxiv against the prior work, both quantitatively and qualitatively. GDPR Enforcement Tracker exclusively focuses on DPA decisions with financial penalties; it contains 1428 such documents. GDPRhub expands this scope to include court decisions in addition to financial penalties; it contains 1594 documents. In contrast, GDPRxiv comprises of 3343 precedents and 5600 guidances (i.e., cumulatively 5× more than the prior work). However, even if we focus only on the precedents, GDPRxiv is 2× bigger than the prior work. This is primarily because GDPRxiv includes enforcement decisions where (i) the controller was not found to have violated GDPR, (ii) the controller was issued a warning, criticism, or a reprimand, or (iii) the violation simply did not merit a financial penalty. As we argued in Section-3.2, all kinds of enforcement decisions—whether upheld or dismissed—do evolve the SOTA.

Next, to assess the quality of our precedent documents, we performed a fine-grained, case-level comparison between GDPRxiv and Enforcement Tracker. Figure-3 shows the per-country break up of these two repositories. While GDPRxiv had more data for all but two countries, we found 86 cases that were present in Enforcement Tracker but missing in GDPRxiv. Upon investigation, we identified these to be either from non-official sources or emanating from links that no longer work. In light of these findings, we conclude that GDPRxiv has a *broader scope* than any of the prior work.

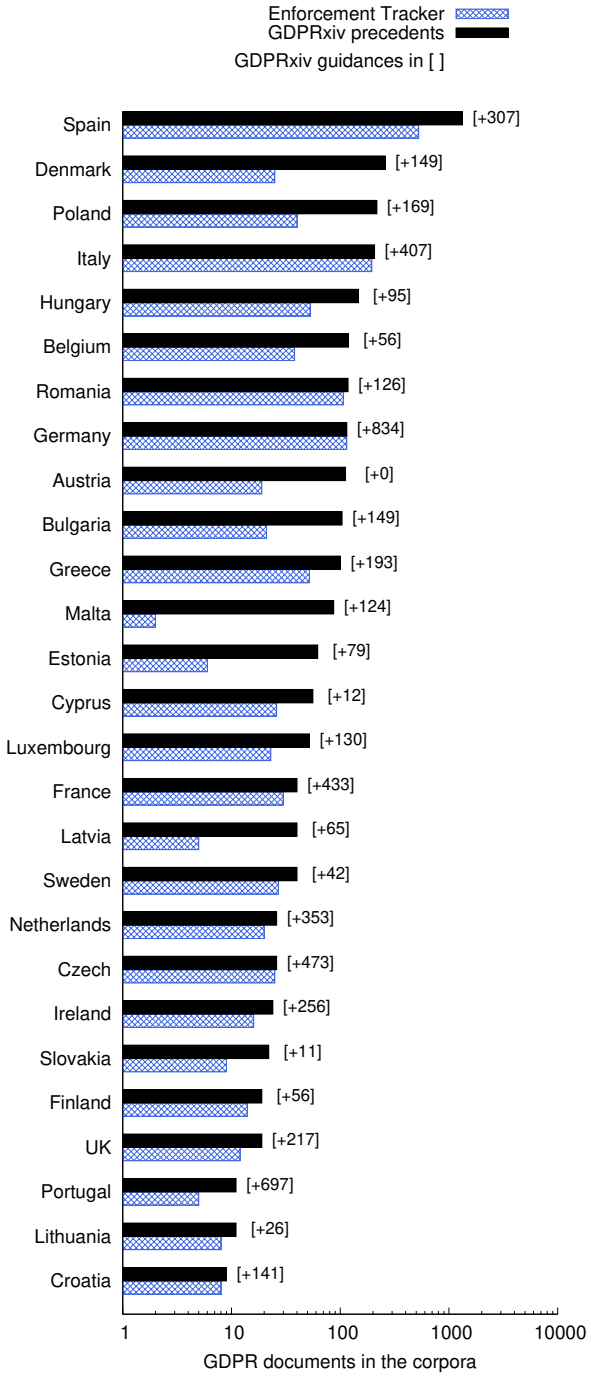
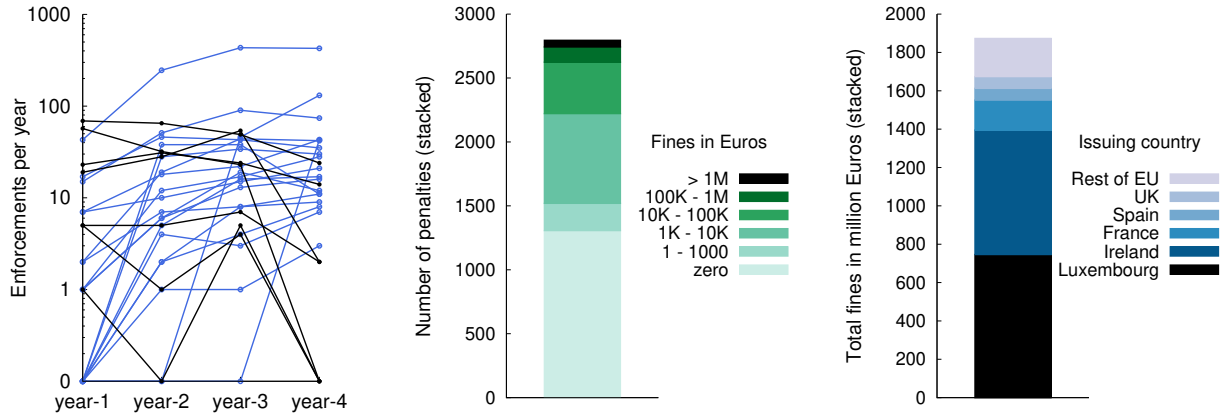


Figure 3: Comparing GDPRxiv's corpus against prior work. GDPRxiv consists of 3343 official enforcement documents, which is 2× more than GDPR Enforcement Tracker and GDPRhub. When official guidances are included, GDPRxiv size expands to 8943, which is 5× more than the prior work.



(a) Change in enforcements over time **(b) Grouping of enforcements by fine amount** **(c) Fines issued by different countries**

Figure 4: Characterizing how GDPR enforcements vary over time, by the amount of fine, and across countries

5.2 Enforcement Trends and Insights

To get a deeper understanding of GDPR's enforcement, we analyze the corpora at the metadata level (i.e., enforcement date, DPA, amount of penalty, and cited articles). This helps us characterize how the enforcements are evolving over the years and across the countries, as well as identify what the focus areas of enforcements are. To do so, we consider all the precedents in GDPRxiv dated between 25-May-2018 and 24-Nov-2022. In this period, there were a total of 3343 enforcement decisions, which translate to 2.03 enforcements every day on average.

Enforcements over time. Figure-4a shows how enforcements evolving over the years. The X-axis marks the four full years of GDPR and the Y-axis measures the number of enforcements brought forward in a given year. Each line in the graph represents a DPA. If a DPA has more enforcements in each of the years 2, 3, and 4 compared to year-1, then we plot it in blue, and we plot all other DPAs in black. We see that two-thirds of the DPAs have expanded on their GDPR activities over the years. In raw numbers, year-1 saw a total of 275 enforcements, year 2 through 4 had 663, 1021, and 1012 enforcements respectively (resulting in a 2.7× increase over four years).

Enforcements and fine amount. GDPR grants DPAs broad authority in levying financial penalties on data controllers that fail to comply with GDPR. While $\mathscr{G}83$ 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). Figure-4b groups all the enforcement decisions[5] into six distinct penalty bins. We see that 46% of enforcements carried no financial penalties at all, and in 80% of the cases, the penalties did not exceed €10K. Million Euro penalties were levied on <2% of violators. This is a reflection of GDPR's *proportional penalty*, wherein the financial penalty is determined by not just the infringement but also the scale and scope

[5]we exclude the cases that were dismissed, rejected, refused, or found to have no infractions from this graph. This brings the total enforcements from 3343 to 2795.

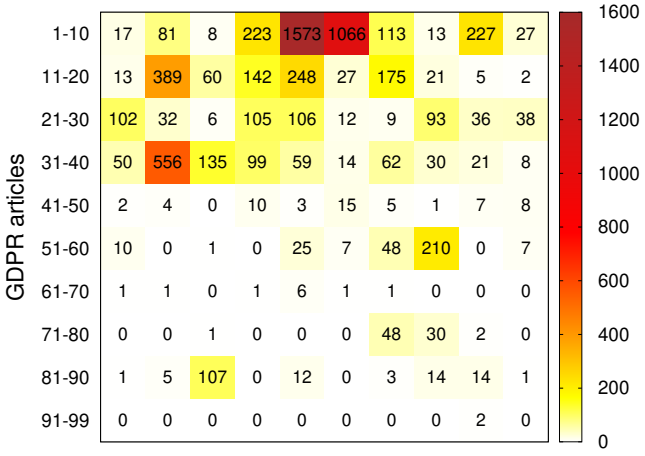


Figure 5: Frequency of citation for each article of GDPR

of its impact (which in turn, puts large companies holding high volumes of personal data at an elevated risk for receiving oversized penalties).

Enforcements across countries. Figure-4c tabulates the fines issued by each of the EU countries over the last 4.5 years. We sort the countries by their total fine amount and stack them to represent the EU-wide total of €1.87B. We see a heavy skew with the top-3 countries accounting for more than 80% of all GDPR fines. In fact, the bottom half of the countries have collectively issued a fine of €12.5M, which is <1% of the EU-wide total.

If we shift our focus to the total number of enforcements as opposed to the amount of fine, we see a different skew. Figure-3 plots the number of enforcements along the X-axis and the issuing countries along the Y-axis. As we see, the top-3 DPAs are responsible for 54% of all enforcements, while the bottom half only add up to 10%. An interesting observation is the lack of correlation between DPAs issuing large number of enforcements versus those levying

No	GDPR article and its key clauses	What they regulate (paraphrased)
$\mathscr{G}5$	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
$\mathscr{G}6$	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; etc;
	(2) Data usage beyond initial purpose	Four conditions including establishing a link between the two purposes; analyzing the consequences of new purpose; etc;
$\mathscr{G}32$	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 the category of data

Table 3: Key articles of GDPR that represent the focus areas of enforcements

heavy fines. In fact, there is only one country (Spain) common between the top-5 heavy enforcers and top-5 heavy finers.

Articles cited in enforcements. The articles of GDPR, which are 99 in number, could be grouped into five broad categories. $\mathscr{G}1$-11 contain definitions and principles of personal data processing; $\mathscr{G}12$-23 establish the rights of the people; $\mathscr{G}24$-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-5 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 ($\mathscr{G}$33-34) or prominent rights such as *Right To be Forgotten* ($\mathscr{G}17$), *Right to Object* ($\mathscr{G}21$), and *Right of Access* ($\mathscr{G}15$) have not resulted in a significant number of citations.

Given the importance of articles 5, 6, and 32, we would be remiss not to have a discussion on them. Table-3 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 regulators are placing on sound data management practices starting from how personal data is to be procured ($\mathscr{G}5$), how it is to be processed ($\mathscr{G}6$), and how the security infrastructure is to be designed and operated ($\mathscr{G}32$). 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.

5.3 Reducing Uncertainty in Compliance

As described in Section-2.2, the computing community faces uncertainties in various stages of designing and operating compliant systems. Our goal here is to demonstrate the utility of GDPRxiv in reducing this uncertainty. To keep this evaluation concrete, we focus on one of the prominent articles of GDPR, *$\mathscr{G}17$: Right to be Forgotten* (RTBF). Then, we ask the question: does the enforcement corpora help us identify common RTBF failures as seen in the field? If so, do these translate into actionable insights for the computing community?

Methodology. To answer these questions, we gather all the enforcement documents that cite $\mathscr{G}17$ (a total of 175 decisions in the first 4.5 years), and then analyze them from a computing perspective. This task was carried out by the lead author, who is a CS graduate student but had previously taken a course on privacy regulations from the law college. For each document, they extracted the main technical/policy reason(s) that led to the enforcement. We followed an open coding methodology i.e., we did not start with any known list of failures, but instead extracted the failures organically from each document. While a vast majority of the documents contained a single reason for failure, 16% of them had two or more reasons. To validate the correctness, a team of three authors (two with CS background and the other, a law scholar) randomly selected 25 documents and generated their failure reasons. These were then compared against that of the lead author. We did not find any discrepancies in coding. Next, we observe that these failures could be organized into five logical groups, each of which represent a

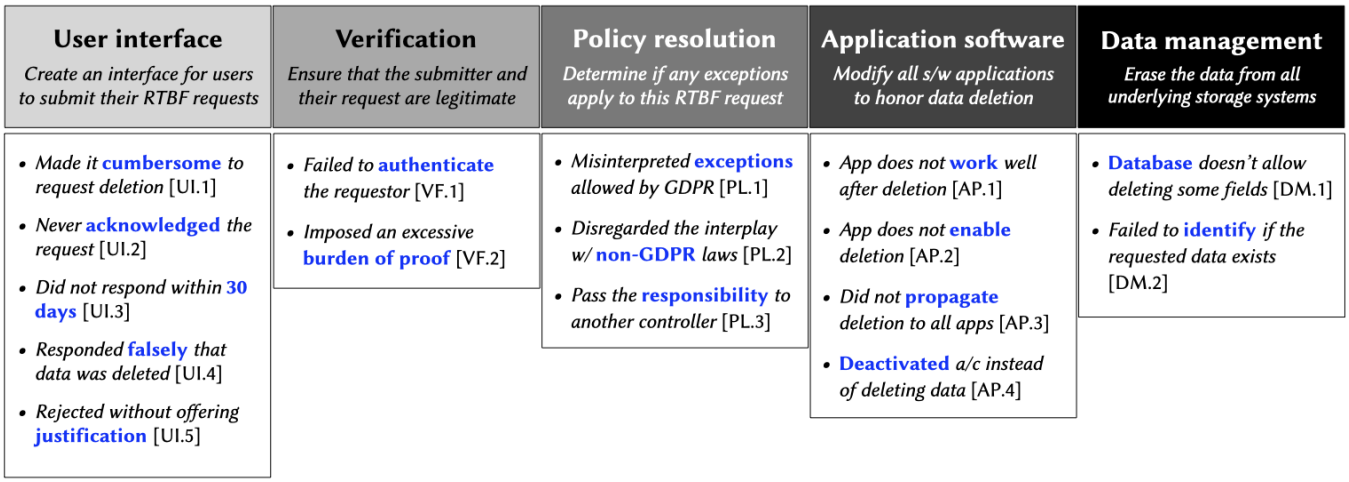


Figure 6: Organizing the RTBF failures into five broad categories. Appendix A lists all the decisions within each category.

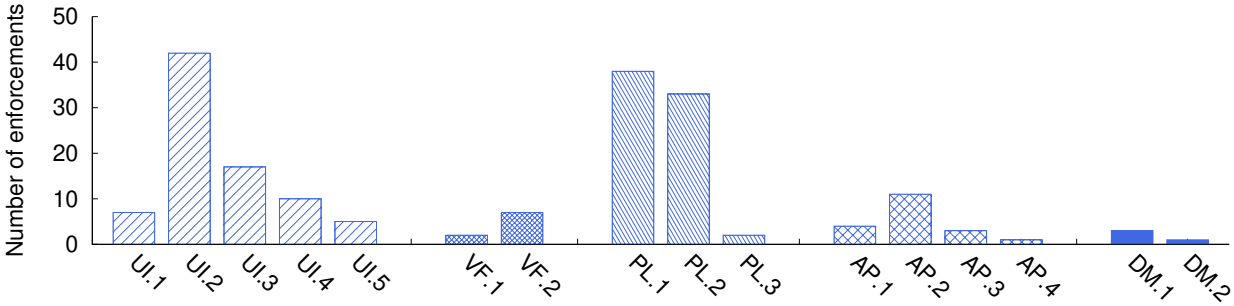


Figure 7: Number of enforcements received by each of the failure categories.

high-level functionality to be supported by an RTBF-capable system. Finally, we study the frequency distribution of these failures to help identify areas that are most vulnerable. Section-5.3.3 explains the scope and limitations of our approach.

5.3.1 Failures in Implementing the RTBF Capability.

At its core, $\mathscr{G}17$ states that *the data subject shall have the right to obtain from the controller the erasure of personal data concerning him or her without undue delay*. It lays out seven conditions, one of which must be met in order to request the deletion; and five exceptions, which if applicable would invalidate the RTBF requests. Figure-6 shows a grouping of the RTBF failures under five broad RTBF functionalities. We must emphasize that this organization is not meant to be a guide for implementing RTBF, but instead to make it easy to study the RTBF failures.

- **User Interface.** The most visible of the functionalities is the interface through which data subjects submit their RTBF requests. For example, Google search provides a webform to submit the URLs to be deleted from their search results [49]. GDPR does not stipulate this to be in electronic format, so organizations could use other interfaces like telephone, mail, etc. Figure-6 lists five common UI failures. While the first four are straight forward, the last one needs elaboration. As an example, consider Spotify, which retains credit card information of people signing up for free trials (to prevent an abuse of the program). When a user exercised RTBF on that data, Spotify refused to honor it. The Swedish regulator, while agreeing that Spotify had a legitimate interest to retain that data, issued a reprimand for not providing a clear explanation to the user.
- **Verification.** While prior work [43] has chronicled lax verification measures at the onset of GDPR in 2018, we do not see any enforcement decision specific to that problem (which is not too surprisingly since people may not have realized that their data was stolen/deleted by impersonators). Instead, what we see are the companies being penalized for requesting excessive proofs from users seeking RTBF. An example of this is when the Irish regulator fined Groupon for requiring national ID from users for exercising RTBF while requiring no such IDs during account creation.
- **Policy Resolution.** Not all RTBF requests need to be (or should be) honored. Exceptions can come from within $\mathscr{G}17$, from other GDPR articles, and even laws outside of GDPR. The complexity of this component is evident in Google Search's

RTBF experience [13], where it continues to be a manual process with a mean time to resolution of 6 days. Despite these guardrails, in July 2020, the Belgian regulators found Google to be misinterpreting the exceptions when they refused to delist four URLs. Since the URLs were more than a decade old, the regulators decided that $\mathscr{G}6$, which makes the processing lawful, no longer applied to this data. Interestingly, we do not see any cases where the controller was penalized for honoring the RTBF request where they should not have. Another failure category is when a controller deflects their obligation to another controller. For instance, an Austrian content aggregation service refused to honor an RTBF request since the requested content originated on LinkedIn website. They asked the user to contact LinkedIn instead. The Australian regulator fined them since they had to delete that information in their service, independent of whether or not the user reaches out to LinkedIn or how LinkedIn responds to that request.

- **Application Software.** The complexity of this task depends on the software architecture and ecosystem within the organization. For most applications, this may involve recomputing the internal data structures, cleaning up runtime engines and caches, and percolating the deletions to underlying storage systems. RTBF failures at the application level can come in various forms. Swedish regulator reprimanded Rebtel Networks for continuing to send emails even after a customer requested to turn off all email communications, a failure on the part of their applications to enable full deletions. Next, organizations must also remember that when the data is shared between multiple applications or externally with a third-party, they need to have the capability to delete. This bore out when the Danish regulator issued a criticism against Høje-Taastrup municipality for using a third-party processor, who could not delete data on demand. Lastly, applications have to ensure that delete means delete. A case in point: the French regulators fined Brico Prive, a retailer, who was simply deactivating user accounts upon receiving RTBF requests instead of actually deleting their data.
- **Data Management.** If an RTBF request is honored, then the controller has no legitimate reason to retain that data in its storage systems. This may pose significant challenges to organizations that lack good data management practices. For example, the Danish regulators fined Taxa (a Copenhagen-based Taxi company), when they revealed that customer telephone numbers are used as primary keys in their database system, and thus could not be deleted. Another example of bad data management is Clearview AI, which was fined by the UK regulator for their inability to recognize all the photos that belong to a given person (because they had not tagged the photos at the time of collection).

5.3.2 Distribution of RTBF failures.

While our analysis identifies 16 common RTBF failures, it is prudent to understand their frequency distribution. Figure-7 shows the failure categories on the X-axis and their frequency on the Y-axis (interested readers may refer to Appendix A for a full list of cases belonging to each category). First off, we see that UI failures are the most common followed by the policy ones, and they account for 80% of all failures. This is likely because we are in the early days of RTBF, and the UI and policy violations are easier to catch than say, application and data management ones. This also bears a useful takeaway for the computing community: yes, applications and database systems need upgrades for RTBF, but it is more urgent to invest in designing a good UI/UX system, and developing a robust policy for handling RTBF exceptions.

5.3.3 Scope and limitations.

Our analysis takes a *via negativa* approach (i.e., identifying negative results based on real-world enforcements). Consequently, our findings are not a recipe for building compliant systems, but instead useful in weeding out common failures. Second, this list of failures should not be considered exhaustive nor treated as prescriptive (i.e., you may eliminate all the failures that we listed, but still may end up violating RTBF). Finally, we cannot provide clarity on those aspects of RTBF where there are no precedents or guidances from the regulators yet. This includes aspects such as *latency of deletion* (i.e., how soon after the RTBF request, should the data be deleted), *guaranteed deletion* (i.e., should the controller offer a proof that the data was permanently deleted), and how RTBF applies to data used in training AI/ML systems.

6 RELATED WORK

Tracking the enforcement of GDPR and understanding its implications for the computing community is a research area with broad scope. Our work is closely related to GDPRhub [48] and GDPR Enforcement Tracker [37], both of which track GDPR enforcement in the wild. Section-2.3 describes their shortcomings and our methodological advantages in more detail. Next, given the significance of GDPR penalties, researchers have developed models to understand [46] and to predict GDPR fines [45]. Our work enables such research by providing a reliable source of ground truth about GDPR enforcement.

Orthogonal to our focus, some recent work has explored automatic checking of systems for GDPR compliance [28, 36, 53] and generating privacy policies to comply with GDPR [39, 56, 57]. Similarly, there are efforts to enhance the legacy software systems to be GDPR capable [9, 50]. By establishing a SOTA on GDPR enforcement, our work helps ground these efforts in reality.

Finally, similar to our analysis in Section-5.3, many researchers have explored RTBF. These include experience reports from Google Search [13] and Microsoft Bing [41]; understanding the challenges in exercising RTBF from a user perspective [29, 51]; and empirical analysis of RTBF practices in websites [30, 42]. Researchers from Facebook [22] and Boston University [12] have built data management systems that natively support guaranteed deletion in order to meet RTBF requirements. In contrast to all these work, we approach RTBF from an enforcement perspective, and shed light on common RTBF failures.

7 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.

An educational tool for data rights. We envision GDPRxiv being adapted as a pedagogical tool for data-driven education and exploration of GDPR. 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 [25], Google cluster traces [44], and Million privacy policies [10] that have been widely used in both academic settings and research environments.

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 criterion (say CCTV, Oracle DB, or article-13) gets posted in the SOTA repository. One way to accomplish this is to enable programmatic access to GDPRxiv via REST APIs. We intend to approach this from first principles: by engaging with different users of our platform—programmers, enforcement agencies, policy makers, and educators to understand their use case and to evolve GDPRxiv accordingly.

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.

8 CONCLUSION

In this work, we make an argument that having a well understood SOTA is paramount for the computing community to comply with novel data regulations like GDPR. We define what the SOTA for GDPR is, 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.

ACKNOWLEDGMENTS

This research is supported by University of Iowa's faculty startup grant.

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] 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).
[3] 2018. California Consumer Privacy Act. *California Civil Code, Section 1798.100* (Jun 28 2018).
[4] 2018. Data Deletion on Google Cloud Platform. https://cloud.google.com/security/deletion/.
[5] 2018. General Law for the Protection of Personal Data (LGPD). *Brazil statutory law 13.709* (Aug 14 2018).
[6] 2020. Data Security Trends 2020-21 Q1. https://ico.org.uk/action-weve-taken/data-security-incident-trends/.
[7] 2021. Consumer Data Protection Act. *Virginia Acts of Assembly, 2021 Special Session I* (Mar 2 2021).
[8] 2021. Cybersecurity Conferences 2021 - 2022. https://infosec-conferences.com/.
[9] Archita Agarwal, Marilyn George, Aaron Jeyaraj, and Malte Schwarzkopf. 2021. Retrofitting GDPR compliance onto legacy databases. *Proceedings of the VLDB Endowment* 15, 4 (2021), 958–970.
[10] 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*.
[11] 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.
[12] Manos Athanassoulis, Subhadeep Sarkar, Zichen Zhu, and Dimitris Staratzis. 2022. Building deletion-compliant data systems. *A Quarterly bulletin of the Computer Society of the IEEE Technical Committee on Data Engineering* 45, 1 (2022).
[13] Theo Bertram, Elie Bursztein, Stephanie Caro, Hubert Chao, Rutledge Chin Feman, Peter Fleischer, Albin Gustafsson, Jess Hemerly, Chris Hibbert, Luca Invernizzi, et al. 2019. Five years of the right to be forgotten. In *ACM CCS*.
[14] 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
[15] European Data Protection Board. 2021. Our Documents. https://edpb.europa.eu/about-edpb/about-edpb/who-we-are_en.
[16] 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.
[17] Anu Bradford. 2020. *The Brussels effect: How the European Union rules the world*. Oxford University Press, USA.
[18] Julie Brill. 2018. Microsoft's commitment to GDPR, privacy and putting customers in control of their own data. In *Microsoft Blog*.
[19] Sergey Brin and Lawrence Page. 1998. The Anatomy of a Large-Scale Hypertextual Web Search Engine. In *ACM WWW*.
[20] Ilias Chalkidis, Manos Fergadiotis, Prodromos Malakasiotis, Nikolaos Aletras, and Ion Androutsopoulos. 2020. LEGAL-BERT: The Muppets straight out of Law School. In *Findings of the Association for Computational Linguistics: EMNLP 2020*. 2898–2904.
[21] Aloni Cohen and Kobbi Nissim. 2020. Towards formalizing the GDPR's notion of singling out. *PNAS* 117, 15 (2020), 8344–8352.
[22] Katriel Cohn-Gordon, Georgios Damaskinos, Divino Neto, Joshi Cordova, Benoît Reitz, Benjamin Strahs, Daniel Obenshain, Paul Pearce, and Ioannis Papagiannis. 2020. DELF: safeguarding deletion correctness in online social networks. In *Proceedings of the 29th USENIX Conference on Security Symposium*. 1057–1074.
[23] 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*.
[24] Brian Cooper, Adam Silberstein, Erwin Tam, Raghu Ramakrishnan, and Russell Sears. 2010. Benchmarking Cloud Serving Systems with YCSB. In *ACM SoCC*.
[25] 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*.
[26] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1*. 4171–4186.
[27] 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/.
[28] Danny S Guamán, Jose M Del Alamo, and Julio C Caiza. 2021. GDPR compliance assessment for cross-border personal data transfers in android apps. *IEEE Access* 9 (2021).
[29] Hana Habib, Sarah Pearman, Jiamin Wang, Yixin Zou, Alessandro Acquisti, Lorrie Faith Cranor, Norman Sadeh, and Florian Schaub. 2020. " It's a scavenger hunt": Usability of Websites' Opt-Out and Data Deletion Choices. In *Proceedings*

of the 2020 CHI Conference on Human Factors in Computing Systems. 1–12.
[30] Hana Habib, Yixin Zou, Aditi Jannu, Neha Sridhar, Chelse Swoopes, Alessandro Acquisti, Lorrie Faith Cranor, Norman Sadeh, and Florian Schaub. 2019. An empirical analysis of data deletion and opt-out choices on 150 websites. In *Proceedings of the 15th Symposium on Usable Privacy and Security (SOUPS 2019)*. 387–406.
[31] Hamza Harkous, Kassem Fawaz, Rémi Lebret, Florian Schaub, Kang G Shin, and Karl Aberer. 2018. Polisis: Automated analysis and presentation of privacy policies using deep learning. In *27th USENIX security symposium*. 531–548.
[32] Hamza Harkous, Sai Teja Peddinti, Rishabh Khandelwal, Animesh Srivastava, and Nina Taft. 2022. Hark: A Deep Learning System for Navigating Privacy Feedback at Scale. In *2022 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2469–2486.
[33] Jonathan M Hsieh, Steven D Gribble, and Henry M Levy. 2010. The Architecture and Implementation of an Extensible Web Crawler. In *USENIX NSDI*.
[34] Melvin Johnson, Mike Schuster, Quoc V Le, Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda Viégas, Martin Wattenberg, Greg Corrado, et al. 2017. Google's multilingual neural machine translation system: Enabling zero-shot translation. *Transactions of the Association for Computational Linguistics* 5 (2017), 339–351.
[35] Elaine C. Khoong, Eric Steinbrook, Cortlyn Brown, and Alicia Fernandez. 2019. Assessing the Use of Google Translate for Spanish and Chinese Translations of Emergency Department Discharge Instructions. *JAMA Internal Medicine* 179, 4 (04 2019), 580–582. https://doi.org/10.1001/jamainternmed.2018.7653
[36] Sophia Kununka, Nikolay Mehandjiev, and Pedro Sampaio. 2018. A comparative study of android and IoS mobile applications' data handling practices versus compliance to privacy policy. *Privacy and Identity Management. The Smart Revolution: 12th IFIP WG 9.2, 9.5, 9.6/11.7, 11.6/SIG 9.2*. (2018), 301–313.
[37] CMS Law. 2021. GDPR Enforcement Tracker. https://www.enforcementtracker.com/.
[38] Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. 2020. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics* 36, 4 (2020), 1234–1240.
[39] Thomas Linden, Rishabh Khandelwal, Hamza Harkous, and Kassem Fawaz. 2020. The privacy policy landscape after the GDPR. *Proceedings on Privacy Enhancing Technologies* 2020, 1 (2020), 47–64.
[40] Xiong Liu, Greg L Hersch, Iya Khalil, and Murthy Devarakonda. 2021. Clinical trial information extraction with BERT. In *2021 IEEE 9th International Conference on Healthcare Informatics (ICHI)*. IEEE, 505–506.
[41] Microsoft. 2022. Right to be forgotten Requests. https://www.microsoft.com/en-us/corporate-responsibility/right-to-be-forgotten-removal-requests-report.
[42] Mohsen Minaei, Mainack Mondal, and Aniket Kate. 2022. Empirical Understanding of Deletion Privacy: Experiences, Expectations, and Measures. In *31st USENIX Security Symposium*. 3415–3432.
[43] James Pavur and Casey Knerr. 2019. Gdparrrrr: Using privacy laws to steal identities. *arXiv preprint arXiv:1912.00731* (2019).
[44] Charles Reiss, John Wilkes, and Joseph L Hellerstein. 2011. Google cluster-usage traces: format+ schema. *Google White Paper* (2011).
[45] Jukka Ruohonen and Kalle Hjerppe. 2020. Predicting the Amount of GDPR Fines. *arXiv preprint arXiv:2003.05151* (2020).
[46] Marlene Saemann, Daniel Theis, Tobias Urban, and Martin Degeling. 2022. Investigating GDPR Fines in the Light of Data Flows. *Proceedings on Privacy Enhancing Technologies* 4 (2022), 314–331.
[47] 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.
[48] Max Schrems. 2021. GDPRhub. https://gdprhub.eu/.
[49] Google Search. 2022. Personal Data Removal Request Form. https://reportcontent.google.com/forms/rtbf.
[50] Supreeth Shastri, Vinay Banakar, Melissa Wasserman, Arun Kumar, and Vijay Chidambaram. 2020. Understanding and Benchmarking the Impact of GDPR on Database Systems. *Proceedings of the VLDB Endowment* 13, 7 (2020), 1064–1077.
[51] Kejsi Take, Kevin Gallagher, Andrea Forte, Damon McCoy, and Rachel Greenstadt. 2022. "it feels like whack-a-mole": User experiences of data removal from people search websites. *Proceedings on Privacy Enhancing Technologies* 3 (2022), 159–178.
[52] Adam Tanner. 2007. Google seeks world of instant translations. In *Reuters*. https://www.reuters.com/article/us-google-translate-idUSN1921881520070328.
[53] 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*.
[54] Nicole Wetsman. 2021. Google Translate still isn't good enough for medical instructions. In *The Verge*.
[55] Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, et al. 2016. Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144* (2016).
[56] Le Yu, Tao Zhang, Xiapu Luo, and Lei Xue. 2015. Autoppg: Towards automatic generation of privacy policy for android applications. In *Proceedings of the 5th Annual ACM CCS Workshop on Security and Privacy in Smartphones and Mobile Devices*. 39–50.
[57] Sebastian Zimmeck, Rafael Goldstein, and David Baraka. 2021. PrivacyFlash Pro: Automating Privacy Policy Generation for Mobile Apps.. In *NDSS*.

A RTBF-RELATED ENFORCEMENTS

Following is the list of enforcement decisions that informed our set of common RTBF errors (shown in Figure-6). They are identified by their location in the GDPRxiv repository.

[UI.1] Made it cumbersome to request deletion

czech_republic/inspections/2020 - 1st/41ed57a47eaf11f21c8aebf905bac4c3
czech_republic/inspections/2021 - 2nd/5070ebcb783df31008e4a1a7d2448b01
france/decisions/d7d0b8a221b124b427472b38837b97d8
italy/injunctions/335fdc1ec98c8cfc929935422a54dcfa
poland/decisions/ed1a16b5edd8f468a17ddf95aa75bfc9
hungary/decisions/cb0adc9596b5f7cd784c248531a9ca5f
spain/decisions/8422be38401c39278d6f54cb9d76af3a

[UI.2] Never acknowledged the request

belgium/decisions/69aa55ca7afa3fe5bb3eab10eb8cb591
belgium/decisions/b4f1beee9a7e34686560b4293fef6f43
belgium/decisions/d8012162ed856fc95f4f8f2cbc518734
belgium/decisions/f2dd4ff35314a7f8767bcae4b51781b1
belgium/decisions/d7f15b2923b66c9cd7be595f105a7707
belgium/decisions/9c58ecf411c1143ab423318e643ab544
belgium/decisions/9c3f747cfb0e65d25eb1be2bbb649465
belgium/decisions/6369fd2ad04dc582b5c0a44ea5d7c458
czech_republic/decision of president/544392c8c19a353c154fc5dd11d30464
czech_republic/decision of president/7e91a38841dc8112b58c91465d172ecd
czech_republic/inspections/2021 - 2nd/5070ebcb783df31008e4a1a7d2448b01
estonia/prescriptions/3bd02055af6a19d3f8d991259608ef6b
estonia/prescriptions/26b006d247bff40fe67b6f7f53f3e672
estonia/prescriptions/80cb7e0aea50a4f06ee2da6441d3669f
estonia/prescriptions/602c0c5004017ca7b52c696a45651171
estonia/prescriptions/e2050e9886368715cb760bbca1e4b20a
finland/2020/264f932ecdbea7dd7ebbf9fe5bef867f
finland/2021/892210c9133bb609afc85f27a7b4091b
finland/2022/3ec8ab336b49a6ee6bd52ce97100bbd0
finland/2022/8f3e27384bdf25553aa51ef57be3c0cc
greece/decisions/2e577b2c95b7f7477e83c714dd4bc2bf
greece/decisions/63fc0f397dd505630ef92945043a1d95
greece/decisions/b4b7ca9b7fdc9f3e1baeed82104e2b1f
greece/decisions/7c23a5f208afafc4c562d46afa59fba0
hungary/decisions/c79a1dcf9c11fecebd0d5ddcd1968b12
hungary/decisions/7f581067e6bbf9d2be935274fe482605
hungary/decisions/cb0adc9596b5f7cd784c248531a9ca5f
italy/injunctions/636ed038596189c2d6dff0e0176aebd0
italy/injunctions/f4b1ca3332a217d91f8985eb3440364f
italy/injunctions/38d5602932a7add5b6fd7f548b56bc22
italy/injunctions/a2e1c2e3c83908ad5db5b1172c2d1d30
italy/injunctions/7f94061a5c084a3bcd208c09ce8dac7f
italy/injunctions/dbc7b34762574caddfaf512709b217d1
poland/decisions/8bb73445e75a46b140d3e7ebba5f0873
poland/decisions/efa62d856b17735c98ea2dac9d8ff0ff
spain/decisions/eb09cb7a8e2e8ce5b520c5b67588a439
spain/decisions/992bc929d0b7fc563cda2c9988c88f77
spain/decisions/b87870baea36e5abbfe1d7ea327d33fc
spain/decisions/cbdf7490b67bf38ad5dcbf055a5e46b1
spain/decisions/546a0177091f33f6e030e4a2860f97f9
spain/decisions/369ad3ae8fe156ce8187a9847bfb83aa

spain/decisions/0d5585313022e88be1fa42a002bde8f9

[UI.3] Did not respond within 30 days

belgium/decisions/d99aca55b0784fbf1fc762db6fdfa365
belgium/decisions/88a9d746a7c14dc960ea8c572cac3733
cyprus/decisions/50c1742de42dc6696fad79203382614d
czech_republic/decision of president/f0b2314e8e1097314b3b6c3a2a175497
denmark/decisions/dba1a7d5856c4f385abb3d396c62b3ac
denmark/decisions/dba1a7d5856c4f385abb3d396c62b3ac-05-07-2019
greece/decisions/c9e39378bbd098741af7111a3a8b0c6
hungary/decisions/0010be447dd92b37ea8fa471cbb2d490
hungary/decisions/44eb563c23c0e50f5e13ba7f6a9e69ea
ireland/decisions/5839b972f8591f7cc59873ce8cad46bb
italy/injunctions/63329e306844908466fbfe73e472c198
romania/decisions/3d2f2d3fbc3b04104a8e2e3244cccbef
romania/decisions/e3684cd3b38168434db879879889ee32
spain/decisions/21ad28399878f812487ccd4b7348cd23
sweden/decisions/8b36e9207c24c76e6719268e49201d94
sweden/decisions/8e6c327b42cb1b718cbacd2fea4ccc01
sweden/decisions/6885ac64b61d589b263f0b13a325b29f

[UI.4] Responded falsely that data was deleted

belgium/decisions/d7e5029e11da77bc33b582191c61e062
finland/2020/264f932ecdbea7dd7ebbf9fe5bef867f
france/decisions/f0d53f21bbad3439ad931482e7f3e4b0
greece/decisions/02c3803e47e2bb24a2feb07e1eb0a1ea
italy/decisions/335fdc1ec98c8cfc929935422a54dcfa
spain/decisions/5e5681ee8d244614c8bae2757002c1d2
spain/decisions/8d228045371fe7e9904d1aa560377ed7
spain/decisions/9ffbbf43703e04245e01b43041b27aef
spain/decisions/6e6fee4a4b7992d8dcf2155926053303
sweden/decisions/2409a2f6e059321f307932d29d84970b

[UI.5] Rejected without offering justification

austria/decisions/207d8557cfa21947518e725aaa94e9b5
belgium/decisions/0b7e8ba2621cde6a7ed71ee3b602c7c1
belgium/decisions/c21ffbed09c1cd6af3e3252914c2c7b9
hungary/decisions/15035a601b489d139753ee651012d265
sweden/decisions/7cab72c2addfac700602cfd09ad1d3fc

[VF.1] Failed to authenticate the requestor

denmark/decisions/2325316e19e277ce201bdd7b26d64ba8
uk/enforcements/36d3493428d2946728fa85845f4d9605

[VF.2] Imposed an excessive burden of proof

austria/decisions/6e747ce53e10256e680114d361330f03
denmark/decisions/43bed9699a610c4d50b782e2b42c6f83
finland/2021/892210c9133bb609afc85f27a7b4091b
finland/2022/3ec8ab336b49a6ee6bd52ce97100bbd0
ireland/decisions/5120d05d4129907e1b934b23a6ad8ecc
ireland/decisions/5839b972f8591f7cc59873ce8cad46bb
sweden/decisions/6885ac64b61d589b263f0b13a325b29f

[PL.1] Misinterpreted exceptions allowed by GDPR

austria/decisions/047eb2831ade8fad73666feda6f88161
austria/decisions/a47f66a0c773e8344f104cd486b77322
austria/decisions/ad2f701244f28c072e0ea63a282876c5
austria/decisions/ebfad367c6eaf8e8efe12bd2a718b814
austria/decisions/f4b6ae96b14b80ba5161d28fdcc0166c
austria/decisions/f92c4cf16a9a3ea0c140985f5f541ef3
austria/decisions/fb716f35a3e9313efb43f3d0768eaed0
belgium/decisions/0b7e8ba2621cde6a7ed71ee3b602c7c1
belgium/decisions/5b83dc831011088371b46e98126f6cd6
belgium/decisions/b4f1beee9a7e34686560b4293fef6f43
belgium/decisions/f2dd4ff35314a7f8767bcae4b51781b1
belgium/decisions/d2fc442492f1c507207fca33ee95c6c3
belgium/decisions/fed23f71e2ad14cf1aed3cc98fa842b7
belgium/decisions/9c58ecf411c1143ab423318e643ab544
belgium/decisions/c10b0b0216ccb2b2b503fcc670efc2a2
bulgaria/CPDP decisions/53b04cb22de880f8d3aa9097509401f8
czech_republic/inspections/2019 - 2nd/d621a4b8dd7ff25703e435892c1b45b0
estonia/presciptions/1e90a856a3a1c2285d98b44caac3e6c1
greece/decisions/5a7f727bd50a59414178d6a2bee6337b
greece/decisions/63fc0f397dd505630ef92945043a1d95
poland/decisions/c9a0e26d57b1b8640fc263d93e6e2c81
sweden/decisions/7cab72c2addfac700602cfd09ad1d3fc
finland/2020/104adac666b3452cbb73b7a909220b83
denmark/decisions/2325316e19e277ce201bdd7b26d64ba8
denmark/decisions/3dc0ea28b353d99873844e1115f5a4f8
denmark/decisions/471210f912725f9f9eca569af18181a2
denmark/decisions/d528d82db7c1ccad5ad778013a62c000
denmark/decisions/f8656b35e9dd01be7051346f5ca33744
hungary/decisions/9a7a75c10f432569a91d31f9a3af7240
hungary/decisions/3f9fcb61a5f40147790e89afb60597f7
hungary/decisions/1030c2f97326379d5546725c67289f6a
hungary/decisions/44eb563c23c0e50f5e13ba7f6a9e69ea
hungary/decisions/de1ce9aa50b455a77d31b9c190825594
hungary/decisions/c090f2cf3bd373c4e99f361a161c8f0d
hungary/decisions/7f581067e6bbf9d2be935274fe482605
hungary/decisions/0e9f829473171fc3bf7a06f0b6b68333
spain/decisions/5e5681ee8d244614c8bae2757002c1d2
spain/decisions/c9e2e7491fd0c9406434d57ae73aded2

[PL.2] Disregarded the interplay with non-GDPR laws

austria/decisions/1718756f902c341c35bae6954a5c133d
austria/decisions/21382ae12da61892c8d38ab7bcacad83
austria/decisions/2530a449bae0baf358d69bdd1715b864
austria/decisions/2388e60c83b6cf55d2459fd3d224b482
austria/decisions/7b0505982755fbe1197638b02b8f3219
austria/decisions/7bb786e317101b65a0ea2f81c26a5837
austria/decisions/8c2152a5abf4d8246019b8e4e7cae829
austria/decisions/8d707d931588c5f352aa445a19564306
austria/decisions/9cd0ca59bcd62f7cd2475f46c6d30ab7
austria/decisions/a57409d939e8e071a34315f18d1ca7af
bulgaria/SCA decisions/d9ed7937e2f5364c8d278ed846e0938f
cyprus/decisions/66d3e5674279a91d7787d419aebad415
cyprus/decisions/e6c1c317da64efd6dad166f3588ef424
czech_republic/decision of president/7e91a38841dc8112b58c91465d172ecd
denmark/decisions/dba1a7d5856c4f385abb3d396c62b3ac
denmark/decisions/dba1a7d5856c4f385abb3d396c62b3ac-05-07-2019
denmark/decisions/9f7f00ea95e060c6397fd890b5068d2f
denmark/decisions/d528d82db7c1ccad5ad778013a62c000
denmark/decisions/f8656b35e9dd01be7051346f5ca33744
finland/2021/892210c9133bb609afc85f27a7b4091b
greece/decisions/b3a45c5bb3cbc5ae3f2076ef1b7beaf5-13_01_2022
hungary/decisions/1030c2f97326379d5546725c67289f6a
hungary/decisions/7b395a9c0542fed532884fa880e93b4e
hungary/decisions/44eb563c23c0e50f5e13ba7f6a9e69ea
hungary/decisions/d9e6771264e8e5a669ab0fb021db72c0
hungary/decisions/15035a601b489d139753ee651012d265

hungary/decisions/de1ce9aa50b455a77d31b9c190825594
hungary/decisions/c090f2cf3bd373c4e99f361a161c8f0d
hungary/decisions/d765deb9f9f75bc54608c8a9551667de
italy/injunctions/c2b08f2adefdeb2893d409f69d7ea4b5
poland/decisions/c9a0e26d57b1b8640fc263d93e6e2c81
poland/decisions/273e1c7dd08e195f82f7a2f3d567efd3

[PL.3] Pass the responsibility to another controller

austria/decisions/fb716f35a3e9313efb43f3d0768eaed0
italy/injunctions/7f94061a5c084a3bcd208c09ce8dac7f

[AP.1] Application does not work well after deletion

austria/decisions/e959be9410e201c031e802e5812ad5f9
belgium/decisions/d7e5029e11da77bc33b582191c61e062
finland/2022/3ec8ab336b49a6ee6bd52ce97100bbd0
sweden/decisions/2409a2f6e059321f307932d29d84970b

[AP.2] Application does not enable deletion

belgium/decisions/d7e5029e11da77bc33b582191c61e062
denmark/decisions/156ec0918692ec154f32ed772f8633de
denmark/decisions/c541a4189850125c22b12c175777c779
france/decisions/f0d53f21bbad3439ad931482e7f3e4b0
greece/decisions/02c3803e47e2bb24a2feb07e1eb0a1ea
italy/injunctions/335fdc1ec98c8cfc929935422a54dcfa
italy/injunctions/63329e306844908466fbfe73e472c198
poland/decisions/8bb73445e75a46b140d3e7ebba5f0873
spain/decisions/9ffbbf43703e04245e01b43041b27aef
sweden/decisions/2409a2f6e059321f307932d29d84970b
uk/enforcements/36d3493428d2946728fa85845f4d9605

[AP.3] Did not propagate deletion to all applications

belgium/decisions/68877fcad6e1849be2048d7d5bccafac
belgium/decisions/88a9d746a7c14dc960ea8c572cac3733
france/decisions/f0d53f21bbad3439ad931482e7f3e4b0

[AP.4] Deactivated account instead of deleting data

france/decisions/06c3a2d139249241fef14860130e7a28

[DM.1] Database does not allow deleting some fields

austria/decisions/e959be9410e201c031e802e5812ad5f9
denmark/decisions_2/d240c478648f7c3ebc5426b4e306c652
france/decisions/5d50340fee86052409b4e498101593cc

[DM.2] Failed to identify if the requested data exists

uk/enforcements/36d3493428d2946728fa85845f4d9605