The GDPR and the Reuse of Published Court Decisions; Some Pressing Questions, Illustrated by Developments in The Netherlands

Paper presented at the Law Via the Internet Conference, 7-8 November 2023, Vienna.

Marc van Opijnen[1]

Abstract

Because courts are supposed to settle legal disputes, court decisions contain numerous direct and indirect personal data of all people that are (intentionally, unintentionally or professionally) party to the case at hand. As long as these data are processed for the goal they were collected for – the adjudication of conflicts – such processing can considered to be lawful under the EU General Data Protection Regulation. However, questions arise if such decisions, even if they are pseudonymised, are processed further for other goals: for populating a database within the judiciary, publishing them on a public website, making them available for reuse as open data, or having them indexed by commercial search engines. In this contribution I will discuss these forms of processing from the perspective of the GDPR and related EU legislation, and illustrate them with interesting examples from (mainly) The Netherlands.

1 Introduction

It is uncontested that the online publication of court decisions is of the utmost importance for the transparency and scrutability of the judiciary, as well as for the accessibility and knowledge of the law. At the same time it is acknowledged that persons involved in court cases run the risk of having (often intimate) details of their personal lives exposed. To strike a balance between those two values, in Europe enshrined in Art. 6 and 8 of the European Convention on Human Rights (ECHR),[2] and in Art. 47 and 8 of the Charter of Fundamental Rights of the European Union,[3] the General Data Protection Regulation (GDPR)[4] is of the utmost importance.

Although the GDPR is applicable as from 2018, awareness that the GDPR also applies to the publication and reuse of court decisions seems to be dawning just slowly, while at the same time such documents have features that pose interesting challenges from a GDPR perspective. In this paper I will give an overview of different types of processing of the personal data within court decisions for other goals than those for which they were originally collected: the adjudication of legal disputes.

In § 2, I will first characterise court decisions in relation to the GDPR concepts ‘anonymisation’ and ‘pseudonymisation’ and describe the risks of reidentification.

---

[1] Mr. dr. M. van Opijnen is adviser legal informatics at the Publications Office of The Netherlands (KOOP). The research for this paper was finalised on 15 December 2023. All translations of Dutch legal texts are by the author.

[2] Convention for the Protection of Human Rights and Fundamental Freedoms.

[3] Charter of Fundamental Rights of the European Union.

[4] 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/EC (General Data Protection Regulation).

In the sections that follow, I will discuss, by increasing audience of the dissemination: internal reuse within courts (§ 3), publication on a judiciary website (§ 4), making decisions available as open data (§ 5) and allowing search engines to index them (§ 6). In § 7 special attention will be paid to the processing of personal data of judges. Some final remarks will be made in § 8.

While the questions touched upon are global in nature, I will limit myself to the EU legal framework and the (legal and factual) situation in the Netherlands, while I will also touch upon new legislation in Belgium. Throughout the paper, I will use the words 'anonymisation' and 'pseudonymisation' as defined by the GDPR, even when referring to pre-GDPR situations. I will use the terms 'deidentification' (deleting or obscuring personal data) and 'reidentification' (undoing deidentification) in a neutral sense, regardless of whether it pertains to anonymisation or pseudonymisation.

2 The Nature of Personal Data in Court Decisions

In common parlance, 'anonymisation' means 'removing or obscuring personal data', and hence, the term is often also applied to court decisions from which (at least) direct personal data have been removed. Currently, every decision published in the Dutch public database[5] bears the notification: *"This decision has been anonymised in accordance with the Anonymisation Guidelines."*[6] The database itself nor these Anonymisation Guidelines contain any reference to the GDPR. If they would, the decisions probably would have been labeled as being 'pseudonymised'. Since under the GDPR, these two concepts have a specific legal meaning that deviate from daily speech; the difference basically determines whether the GDPR is applicable or not.

The term 'anonymous information' does not appear in the operative part of the GDPR, but is defined in recital 26 as:

> *"[I]nformation which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable."*

The most important difference with pseudonymous information is in the (ir)reversibility of the deidentification process; Art. 4(5) GDPR defines 'pseudonymisation' as:

> *"[T]he processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person."*

To distinguish anonymisation from pseudonymisation the word 'identifiable' plays a crucial role. Recital 26 explains that to establish whether a natural person is identifiable:

> *'[A]ccount should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly. To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments."*

In practice of course, the distinction between anonymous and pseudonymous information is not always that obvious. Basically, two approaches exist (Spindler and Schmechel). The

[5] https://uitspraken.rechtspraak.nl.

[6] https://www.rechtspraak.nl/Uitspraken/Paginas/Anonimiseringsrichtlijnen.aspx. Accessed on 24 November 2023.

absolute (or objective) approach holds that if there is anyone in the world able to (re)identify the deidentified natural person, the data are pseudonymous. On the other hand, the relative (or subjective) approach takes the 'reasonably likely te be used' criterion explicitly into account. The European Data Protection Board (EDPB) still adheres to the 5/2014 Guidelines[7] of the Article 29 Working Party – its predecessor under the Data Protection Directive[8] – which breathes the absolute approach, while the Court of Justice of the European Union (CJEU) seems inclined to move towards the relative approach. In the Breyer case[9] the Court, following the opinion of the Advocate General,[10] held that there are no 'means likely to be used': *"[I]f the identification of the data subject was prohibited by law or practically impossible on the account of the fact that it requires a disproportionate effort in terms of time, cost and man-power (…)."* In SRB / EDPS,[11] the General Court[12] held that it has to be examined: *"Whether the [recipient has] legal means available to it which could in practice enable it to access the additional information necessary to re-identify the [natural person]."* In the recent Scania case[13] the CJEU stressed the requirement of having lawful means at one's disposal to establish whether a natural person is identifiable, and in OC/Commission[14] the General Court held that a person within a deidentified press release could not considered to be identifiable since: *"[I]t would be necessary to mobilize resources that are not reasonably likely to be used by a reader and would certainly require additional time."*

This raises the question whether court decisions from which (at least) direct personal data have been removed, classify as anonymous or pseudonymous.
A first remark to be made is that, in general, the focus of the GDPR, the guidelines of the EDPB,[15] the Open Data Directive (to be discussed in § 5), as well as most case law and academic literature – e.g. (van der Sloot, van Schendel and López) – relate to deidentification techniques within structured datasets or specific, individualisable elements within data exchanges. Court decisions and comparable documents seem to be ignored, while they have characteristics that pose very specific problems. They are long textual documents, lacking machine readable structures and containing some direct personal data and a wealth of all types of indirect personal data, for which often-used deidentification techniques are irrelevant or useless.

In establishing the GDPR status of court decisions, following the absolute approach, they would classify as pseudonymous, even with all direct and indirect personal data removed, since at least the courts themselves would have the means for reidentification.
According to the relative approach, it should be established whether there are means reasonably likely to be used by the controller or by another person to identify directly or indirectly the natural persons, taking into account: *"[A]ll objective factors, such as the costs*

---

[7] Article 29 Data Protection Working Party, Opinion 05/2014 on Anonymisation Techniques.
[8] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data.
[9] CJEU, 19 October 2016, C-582/14 (Breyer), ECLI:EU:C:2016:779.
[10] Advocate General CJEU, 12 May 2016, C-582/14 (Breyer), ECLI:EU:C:2016:339.
[11] General Court, 26 April 2023, T-557/20 (Single Resolution Board (SRB) / European Data Protection Supervisor (EDPS)), ECLI:EU:T:2023:219.
[12] Since it was an interinstitutional affair, the case was decided by the General Court, not on the GDPR, but on Regulation (EU) 2018/1725 of the European Parliament and of the Council of 23 October 2018 on the protection of natural persons with regard to the processing of personal data by the Union institutions, bodies, offices and agencies and on the free movement of such data, and repealing Regulation (EC) No 45/2001 and Decision No 1247/2002/EC, which is substantively comparable to the GDPR.
[13] CJEU, 9 November 2023, C-319/22 (Scania), ECLI:EU:C:2023:837.
[14] General Court, 4 May 2022, T-384/20 (OC/Commission), ECLI:EU:T:2022:273.
[15] See footnote 7.

of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments" (Recital 26 GDPR). It has to be noted that in nearly all case law discussed above, there was a specific recipient of the data of whom it had to be established whether he might have 'the means reasonably likely to be used' at his disposal. This practically means – as formulated by the AG in the Breyer case[16] – that 'another person' is not 'just any one'. But when it comes to documents intentionally disseminated on a public website, 'another person' actually should be interpreted as 'just any one', since anyone does have access.
Also, in e.g. Breyer and SRB/EDPS, the lawfulness of the reidentification attempts are taken into account to establish whether personal data can considered to be anonymous. As will be discussed in § 5, explicit legal provisions are to be introduced in The Netherlands to prohibit reidentification. Nevertheless, I have to concur with (Groos and van Veen) in their observation that: *"Hacking is an illegitimate act in almost all jurisdictions, (…) that does not mean that it might not happen."* Which leads them to the conclusion that 'reasonably likely' has to be understood also in the sense that reidentification by a *"[T]hird party using* illegitimate *means would require a disproportionate effort in terms of time, cost and manpower, so that that risk of reidentification appears in reality to be insignificant."* (Emphasis mvo.)

To establish whether such reidentification risks actually exist, it shouldn't just be discussed, it should actually be tried. Two recent experiments shed light on the success rate of such hacking attempts. In 2019 (Vokinger and Mühlematter) scrutinised a Swiss database with judicial decisions on admission and price increases of medicinal drugs. Instead of personal data the names of the drugs and the manufacturers were pseudonymised, but for the goals of the experiment this is not too relevant. Reading the text of an individual decision, one was not able to discover the name of the drug or the manufacturer. However, after downloading other publicly available datasets, performing some basic search algorithms and then allowing one hour per decision for manual inspection, the researchers managed to re-identify a staggering 84% of the decisions. In another experiment, (Deuber, Keuchen and Christin) took a random sample from published German court decisions and gave second and third grade law students 35 minutes per decision to re-identify all 'attackable strings' therein, which in some way had been deidentified by the publishing court. In their meticulously described research they concluded that 25% of those strings could be reidentified. Sources for reidentification were mostly internet news channels (23%) and personal pages (21%). The research also unveiled that names of natural and legal persons were most vulnerable and that of the various pseudonymisation techniques 'partially-preserving omissions' (like shortening names to initials) are most easy to hack.
Given these results, the researchers' inevitable conclusion was: *"[T]hat our conservative approach satisfies [and even exceeds] any plausible legal concept of 'reasonable effort,' which means that the supposedly protected information is not anonymous in a legal sense."*[17]
Additionally, in choosing the safeguards to be implemented, also account should be taken of current technological developments – especially the combinatory capabilities of large language models –, of the fact that court decisions often contain personal data that enjoy a higher standard of protection under the GPDR, like the special categories of Art. 9 and data relating tot criminal convictions and offences (Art. 10), of the fact that reidentification or accidental disclosure of personal data in court decisions might easily ruin the lives of those involved, as well as of the fact that still many court decisions (unintentionally) are not being pseudonymised properly at all.

[16] See footnote 10.
[17] Loc. cit. par. 8.

Hence, more research into pseudonymisation techniques is needed, which maybe even should lead to certifications for such software (Art. 42 GDPR), in the knowledge that no perfect solutions will exist (Csányi et al.).

3 Internal Reuse

In many jurisdictions, judicial decisions are stored within the judiciary itself, nowadays more often in electronic than in paper format. Official statuses and access regimes of those documents might differ.

Storing personal data in an internal database after they have been processed for the goal they have been collected for – the resolution of the legal conflict that materialised in the decision to be stored – is a type of further processing under Art. 6(4) GDPR, which needs a legal base, consent or a balanced decision by the controller, taking into account the issues mentioned within that paragraph.

In The Netherlands, internal databases that were also accessible for other courts – via CD-ROM – were introduced in the late nineties. In 1997 the Registration Chamber – a predecessor of the current Data Protection Authority (DPA) – delivered an opinion on the processing of personal data in such collections.[18] It basically held that circulation of non-pseudonymised decisions was only allowed between related departments within the courts (e.g. civil chambers), and that all other access (within or from outside the courts) should only be possible after pseudonymisation.

When moving gradually to the intranet (from 2001 onwards), role-based access measures were added to the organisational restrictions in databases that later evolved into what is now known as the 'E-archive'. (van Opijnen 2014). In 2012 though the department-based restriction was dropped, and access to the national E-archive became only role-based.

In 2022 three complaints against three different courts were filed – by one and the same person – with the Procurator General at the Supreme Court (PG) regarding the processing of personal data within the E-archive. Because also critical questions from within the judiciary had reached the PG, he started a more detailed investigation. The request of the PG for an investigation by the Supreme Court was published on 21 September 2023.[19] At the time of writing there had not been a ruling by the Supreme Court yet. I will only discuss the general observations of the PG about the E-archive, not the individual complaints, since they are of secondary importance.

First though, the position of the PG in this procedure has to be briefly explained. Under Art. 51 GDPR each Member State has to provide for one or more independent DPAs, for monitoring the application of the GDPR. However, according to Art. 55(3) they: *"[S]hall not be competent to supervise processing operations of courts acting in their judicial capacity."* This raises at least two questions. First: what is to be understood as 'in their judicial capacity?' And, secondly: how should the supervision of the courts then be organized? The GDPR is silent on the first issue, but on the second issue recital 20 reads: *"It should be possible to entrust supervision of such data processing operations to specific bodies within the judicial system (…), which should, in particular ensure compliance (…), enhance awareness (…) and handle complaints (…)."*

In 2020, a comparative study showed that these two issues are addressed quite differently (Custers et al.). In The Netherlands, the PG is entrusted with the supervision of the general

[18] Registration Chamber, Opinion on case law databases, 11 March 1997, 95.V.113.03.
[19] Procurator General, 21 September 2023, ECLI:NL:PHR:2023:823.

courts,[20] while the three high administrative courts established the Data Protection Commission for Administrative Law Tribunals.[21]
The definition of 'judicial capacity' was subject of the decision of the CJEU of 24 March 2022.[22] In this case, the question at hand was whether the Administrative Jurisdiction Division of the Netherlands Council of State could, on the day of a hearing, grant access to non-pseudonymised court files to a journalist. The data subjects concerned had filed a complaint at the DPA, which declared the complaint inadmissible, due to Art. 55(3) GDPR. The data subjects challenged the decision of the DPA at the district court Midden Nederland,[23] which referred the case to the CJEU for a preliminary reference. The CJEU decided (point 37): *"(…) [I]t must be held that (…) the processing of personal data carried out by courts in the context of their communication policy on cases before them, such as those consisting in the temporary making available to journalists of documents from a court case file in order to enable them to cover it in the media, inter alia, falls outside the competence of the supervisory authority, pursuant to Article 55(3) of that regulation."* Although the preliminary question did not relate to the publication of court decisions, the formulation of the CJEU gives ample room to conclude that also the further processing of personal data in court decisions might fall under 'judicial capacity'.
Interestingly though, the DPA itself seems to be ambiguous in its position. On the one hand, at the district court it defended a broad interpretation of 'judicial capacity': *"Defendant is of the opinion that making court files accessible for journalists serves the openness and transparency of the judiciary and promotes the public trust in the judiciary. Openness can therefore be regarded as a fundamental pillar of the Rule of Law and is inextricably linked to the judicial task."*[24] On the other hand though, on its own website, the DPA claims explicitly that the: *"Processing of personal data by the courts with regard to the publication of decisions and judgments"* is under the competence of the DPA itself.[25]

In his request for an investigation on the E-archive, the PG starts with the assertion, based on the above-cited decision of the CJEU, that the processing of personal data within the E-archive falls under his the competence, leaving the 1997 opinion of the Registration Chamber unmentioned.
Substantively, the PG bases the processing of personal data for judicial tasks *strictu sensu* on Art. 6(1)(e) GDPR: *"Processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller."* He doesn't exclude that processing personal data in the E-archive is already allowed under this provision, but in any case on Art. 6(4) GDPR (further processing) in conjunction with Art. 23(1)(e), (f), (i) and (j) GDPR. The PG acknowledges that debate is possible on the data minimisation requirement of Art. 5(1)(c) GDPR, but he considers the access limiting measures, together with additional measures that were implemented after a data protection impact assessment had been performed, to be sufficient. And while: *"[T]his would not mean that the goal of achieving large-scale anonymisation (or pseudonymisation) would not be commendable*", it is not deemed to be necessary for GDPR compliancy. Tailored pseudonymisation software has been developed within the judiciary, with the advice to use it on all decisions in the E-archive, but no decision had been taken at the time of the PG's request.

---

[20] Regulation on the supervision of the processing of personal data by the courts and the office of the Procurator General at the Supreme Court.
[21] Regulation on the processing of personal data in administrative courts.
[22] CJEU, 24 March 2022, C-245/20 (X, Z / Autoriteit Persoonsgegevens), ECLI:EU:C:2022:216.
[23] District court Midden-Nederland, 26 May 2020, ECLI:NL:RBMNE:2020:2028.
[24] ECLI:NL:RBMNE:2020:2028, 2020, par. 28.
[25] https://autoriteitpersoonsgegevens.nl/themas/politie-en-justitie/politie-bijzondere-opsporing-en-justitie/gebruik-van-persoonsgegevens-door-justitie, accessed on 28 November 2023.

Some observations can be made. First, as can be read in the last quote, either the difference between anonymisation and pseudonymisation is not understood, or deemed to be irrelevant. The two words are often used in an 'or' conjunction as above, without any additional consideration. Secondly, Art. 25 GDPR – data protection by design and by default – is interpreted in such a way that discretionary decision-making power is left to the joint controllers of the E-archive: the Council for the Judiciary and the court presidents. This is quite opposite the way this has recently been regulated in Belgium. In the Act on the establishment of a Central register for court decisions,[26] the governance structure of this register, its contents, its access policies, supervision and data protection safeguards are regulated in great detail. It should be noted that, while in his opinions for the Supreme Court in regular proceedings, the PG often assesses the state of the law in neighboring countries in situations where the law leaves room for multiple interpretations – as is obviously the case here – the PG does not even mention this Belgian act.

4 Publishing on a Public Website

Like building an internal database, publishing court decisions on a public website is a type of further processing of personal data for a purpose other than that for which they have been originally collected. For that reason, in many EU Member States the publication of (a selection of) court decisions have been laid down by law (van Opijnen et al.). In the Netherlands such a legal framework does not exist, although a legislative proposal is currently in drafting stage.[27] I have discussed the need for such a legal framework elsewhere (van Opijnen 2021); for the remainder of this paper, the assumption is that there will be a sufficient legislative base for the publication of pseudonymised court decisions soon.

If pseudonymisation errors are detected, data subjects have the means available to ask for correction (Art 21(2) GDPR). Although it might not always be explicitly invoked, Art. 17(1) GDPR (the right to be forgotten) establishes the right of the data subject to ask for erasure of such unlawfully published personal data. But for the controller it doesn't suffice to just correct those errors on the website, it has to take into account the processing of the unlawfully published personal data by others as well. Databases with court decisions have substantial numbers of visitors. As long as these users download decisions just for personal use, they are not considered to be processing personal data (Art. 2(2)(c) GDPR). However, if people working at i.a. law firms, universities and governmental organisations are manually downloading and storing substantial numbers of decisions, they are not exempt from the GDPR, even the more if they are, in turn, disseminating those decisions to colleagues, students and employees.

Even if those reusers would like to correct pseudonymisation errors, how should they be aware? Art. 19 GDPR puts a notification obligation on the controller in the event of rectifications or erasures, but in case those data have been published on a website, I would argue that Art. 17(2) GDPR is applicable instead:

2) *"Where the controller has made the personal data public and is obliged pursuant to paragraph 1 to erase the personal data, the controller, taking account of available technology and the cost of implementation, shall take reasonable steps, including technical measures, to inform controllers which are processing the personal data that the*

[26] The Act on the establishment of the Central Register for the decisions of the judiciary and concerning the publication of decisions and for the amendment of the assize procedure regarding the recusal of jurors shall apply from 31 December 2023.

[27] In a letter to Parliament the Minister of Justice and Security informed the Parliament to be considering such a legal framework, while documents that were made public following a request under the Open Government Act reveal that the ministry is indeed preparing such an act.

data subject has requested the erasure by such controllers of any links to, or copy or replication of, those personal data."

Since in a situation like this the publishing of pseudonymised data is legitimate, but not the (unintentional) processing of the non-pseudonymised data, I would also argue that Art. 17(2) is not being overruled by Art. 17(3) GDPR. A 'reasonable step' for the judiciary could be to develop a notification system, by which reusers are alerted, e.g. by e-mail, on erasures of personal data in decisions they have downloaded previously. Of course, it will be the responsibility of the reuser to subscribe to such a service and to follow-up on any erasure warnings they might receive. Such a notification system is not meant for reusers that use specific APIs for mass downloads; such APIs (discussed in § 5) often have such functionalities built in. Users that manually download decisions and want to stay informed about pseudonymisation corrections, should not be forced to make use of such complicated facilities, if they exist at all.

5 Dedicated Reuse Facilities

Websites, preferably equipped with user-friendly search interfaces, are very useful for occasional queries, but some prefer to have access to the dataset by other – automated – means, including the option to build and maintain a full copy of the database. For court decisions such reusers include commercial publishers, academic researchers and legal tech companies. EU Member States agreed that the availability of substantial datasets with court decisions is crucial to develop AI applications in the field of justice.[28]

However, making (pseudonymised) personal data available via such an API has to be regarded as a specific and separate type of processing ('dissemination or otherwise making available') under the definition of Art. 4(2) GDPR, serving a purpose which differs from the purpose of making the data available via a normal website. Hence, the lawfulness of this processing has to be established separately under to Art. 6 GDPR.

In the following, I will first discuss the Open Data Directive, as well as the (proposed) Dutch act for its implementation. Then I will discuss the Dutch situation and relevant case law.

The Open Data Directive (ODD)[29] – a recast of the 2003[30] and 2013[31] Directives on the reuse of public sector information – sets the rules for facilitating the reuse of public sector information across the EU. While the ODD strives to make as many public datasets as open as possible, those datasets often do contain personal data, which the GDPR aims to protect. These conflicting interests are hardly addressed within the ODD itself.

Art. 1(4) ODD reads: *"This Directive is without prejudice to Union and national law on the protection of personal data, in particular Regulation (EU) 2016/679 and Directive 2002/58/EC and the corresponding provisions of national law."*

In addition, recital 52 states that the reuse of personal data is only permissible if the principle of purpose limitation of Art. 5(1)(b) GDPR is met, followed by a literal copy of the definition of 'anonymous information' from the earlier cited recital 26 of the GDPR. However, in Art. 2(7) ODD 'anonymisation' is defined as: *"[T]he process of changing documents into anonymous documents which do not relate to an identified or identifiable natural person, or the process of rendering personal data anonymous in such a manner that the data subject is*

---

[28] Council Conclusions 'Access to justice – seizing the opportunities of digitalisation', OJ 2020/C 342 I/01.

[29] Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information.

[30] Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information.

[31] Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the re-use of public sector information.

not or no longer identifiable." It should be appreciated that the distinction between documents containing personal data on the one hand (like court decisions) and (tabular/statistical) personal data on the other hand is introduced here, but 'pseudonymisation' is not mentioned at all, while the term 'anonymisation' is often used in a context where the word 'pseudonymisation' might be used as well or would be even more appropriate. This leaves room for differences in interpretation or for misunderstandings.

In The Netherlands this lacuna is being filled in the current proposal for the Act Implementing the Open Data Directive (AI-ODD)[32] that is set to amend the current Act on the Reuse of Public Sector Data.[33]
After the amendment, Art. 6(3) and 6(4) of the latter are to contain the following provisions to balance the interests of reusers with data protection rights:

3) *"To protect personal data [public bodies] shall at least impose the following conditions on reusers:*
   a) *the data shall not be used for the reidentification of individuals involved;*
   b) *the data shall not be used to undo pseudonymisation as referred to in Article 4, paragraph 4, of the [GDPR] or other safeguards for the protection of personal data;*
   c) *reusers shall establish appropriate safeguards to comply with the previous two clauses, also in the further processing of the data;*
   d) *the previous three clauses do not apply if the processing only takes place for the purpose of conducting research into anonymization or protection techniques, provided that the results of such research are fed back to the [public bodies] from which the data originate;*
   e) *a breach of these conditions shall be reported by the reuser without delay to [public bodies] from which the data originate;*
   f) *if a reuser passes on the data to a third party, they shall mention from which [public bodies] the data originates and also impose the conditions for further reuse as determined in this clause on that third party.*
4) *The first three clauses do not, in principle, preclude the imposition of restrictive conditions insofar as:*
   (…)
   e) *they relate to the rejection of any liability towards the reuser with respect to the documents made available for reuse;*
   f) *they are necessary for the protection of personal data, including conditions necessary to comply with the principle of purpose limitation, as referred to in [Art. 5(1)(b) GDPR]."*

In The Netherlands reuse facilities for the case law database have been available since 2004. In 2013 the initial FTP-connection was replaced by an API (Application Programming Interface) (van Opijnen et al.). The API is freely usable by anyone, without prior registration, contract or pre-imposed conditions regarding the (re)use of (pseudonymised) personal data, except for a general fair use policy.[34] The API offers an option to request all records that have been added or modified since a date(time) that can be configured by the reuser (e.g. the moment of the previous request). This functionality can be regarded as a technical measure to inform reusing controllers that a data subject has requested the erasure of its data (Art. 17(2)

[32] The ODD should have been transposed into national law by 17 July 2021, but this deadline wasn't made by The Netherlands. The 'Act Implementing the Open Data Directive' has been submitted to Parliament, but has not been scheduled yet for debate in plenary.
[33] Wet hergebruik van overheidsinformatie.
[34] https://www.rechtspraak.nl/Uitspraken/Paginas/Open-Data.aspx.

GDPR). It played a minor role in a recent decision of the District Court of Rotterdam,[35] that also looks to have inspired the above-cited Art. 6(4)(e) AI-ODD regarding the liability of the reuser. A data subject was not pseudonymised properly in two verdicts of the District Court of The Hague from June 2017[36] and March 2018.[37] The pseudonymisation errors were corrected in December 2019 and January 2020 respectively, after having been discovered via a Google search by a relative of the data subject. The data subject's compensation claim at the Council for the Judiciary for the unlawful processing of his personal data was settled out of court. However, by using the API, the decisions containing the pseudonymisation errors were replicated on three websites of Magnova, a private company that hosts several legal websites. Via the modified-since functionality of the API, the errors were automatically restored as well; this went completely unnoticed for Magnova until the data subject came forward with a compensation claim. Magnova refused to pay, the data subject went to court and Magnova summoned the State for indemnification.
On the question whether Magnova could be held accountable for revealing the identity of the data subject, the court ruled that Magnova was a controller, responsible for the processing of these personal data, which was unlawful in absence of any of the legitimate grounds in Art. 6(1) GDPR. And then the Court ruled: *"That [Magnova] obtains the judgments it has published from rechtspraak.nl and processes them automatically does not mean, in view of Article 82(3) of the GDPR, that it is in no way responsible for the infringement."*
The Court also rejected the claim for indemnification, in the absence of any written contract or legal provision making the judiciary responsible towards Magnova for the pseudonymisation errors.

If the AI-ODD is to be approved by Parliament without further amendments, the judiciary will be obliged to impose conditions on the reusers of the case law database, at least those described in Art. 6(3) AI-ODD, and probably also more restrictive conditions based on Art. 6(4)(f) AI-ODD. But before such conditions can be formulated, due to the requirements on purpose limitation of Art. 5(1)(b) GDPR, this specific type of processing needs a proper and specific legal base first, preferably in the bill on the publication of court decisions that is currently being drafted. Since such an act basically weakens the rights of data subjects, it has to contain specific provisions as listed in Art. 23(2) GDPR, taking into account the specific risks of publishing court decisions as mentioned in the last paragraph of § 2 above. Preferably, the question should be posed explicitly whether a public reuse API is desirable at all. In the new Belgian act, mass download is forbidden completely, although access can be granted to specific groups, like journalists.
It might be considered whether the Data Governance Act (DGA)[38] can be used to strike a balance between the conflicting interests. The DGA offers a framework for facilitating access to documents containing personal data that fall outside the scope of the ODD if the access regime of those data excludes or restricts access to such data for reasons of data protection. 'Data intermediation services' facilitate the technical implementation of the data access between data holders and data reusers. Hence, the DGA does in itself not create an obligation to share data, but might offer a possibility to make reuse facilities, like an API, available for specific groups of reusers, like legal publishers or research institutes.

[35] District court Rotterdam, 18 November 2022, ECLI:NL:RBROT:2022:9597.
[36] District court The Hague, 17 June 2017, ECLI:NL:RBDHA:2017:6533.
[37] District court The Hague, 9 March 2018, ECLI:NL:RBDHA:2018:2730.
[38] Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act).

6 Exposing to Search Engines

Under Art. 4(2) GDPR, exposing web content to general search engines, like Google, Duckduckgo or Bing, is a type of processing on its own: the audience is enlarged, the impact of possible data breaches is more severe and specific technology is being used. Hence, legislator and controller should ask themselves which legitimate goals could justify such a massive dissemination of personal data, even if they are pseudonymised.
I will first briefly discuss the technical measures for regulating search engines' access to web repositories, followed by a brief analysis of an interesting decision of the Federal Court of Canada and the current state of play in the Netherlands.

To regulate which sections of a website can be indexed by web crawlers (i.e. software to roam the web for information to be indexed by search engines or other data lurkers) the *robots exclusion protocol* is generally used, also known as *robots.txt*.[39] Robots.txt is a small text file placed in the root of a website, informing web crawlers about their access rights. The protocol is a gentlemen's agreement; it is non-enforceable and can be circumvented by malicious web crawlers. Therefore, as an additional layer of protection, many websites implement IP monitoring measures that block requests from an IP address if it surpasses some predefined threshold in the number of requests per time unit. In the context of the GDPR, these measures – robots.txt in combination with IP monitoring – could be used as safeguards for protecting personal data, as in Art. 6(4)(e), Art. 23(2)(d) or 25(1) GDPR.

While robots.txt is basically a protocol to keep web crawlers out, the sitemaps protocol on the other hand is aimed at assisting crawlers to index information on a website faster and more efficiently. Sitemaps are especially of great help if the information to be indexed cannot be found by just following hyperlinks. Sitemaps can also be used to implement measures to inform a search engine of erasures under Art. 17(2) GDPR, e.g. by using the attribute *last modification date* or by producing sitemap files on a daily basis, only containing the inserts and updates of a particular day.[40]
The use of protective measures like robots.txt and IP blocking played an important role in the decision of the Federal Court of Canada, A.T. v. Globe24h.com, 2017 FC 114.[41] The Canadian Legal Information Institute (CanLII) publishes vast amounts of courts decisions, freely accessible for all. Given the open court principle, inherent to common law systems like in Canada, court decisions are published without deidentification, except for specific categories or by decision of the judge. However, to protect those personal data it is prohibited to crawl the CanLII database or to download its contents in bulk. To enforce this, CanLII applies the robots exclusion protocol as well as additional IP monitoring and blocking measures.
The website Global24h.com made a business out of illegally downloading decisions from CanLII in which people of Romanian origin were a party, and republishing those decisions on its own website, without using robots.txt. As a result, such decisions were indexed by Google. If victims were complaining at Global24h.com about their personal data suddenly being exposed via Google, they were extorted to pay considerable amounts to have their names removed from that website. One of the victims complained at the Privacy Commissioner of Canada and the case ended at the Federal Court. Apart from the fact that Global24h.com was found to be illegally processing personal data, the court considered the protective measures taken by CanLII to be sufficient.

---

[39] https://www.robotstxt.org.
[40] As implemented in the ECLI search engine. (van Opijnen and Ivantchev).
[41] https://canlii.ca/t/gx6bl.

The robots.txt on the Dutch case law database[42] is extremely lenient, and to serve web crawlers even better, also a sophisticated sitemap structure has been implemented. As a result, Dutch court decisions are easily found via common search engines, which also explains why many pseudonymisation flaws (like in the Magnova case described above) are detected via Google searches.

7 Personal Data of Judges

As a general rule, in continental Europe, parties to a case are being pseudonymised, not those professionally involved: judges, clerks, lawyers or other legal representatives.
In France, in the aftermath of a research report on judicial biases in asylum cases[43] and based on a comprehensive report (Cadiet), Art. L10 of the Code of Administrative Justice was adapted,[44] not only allowing the pseudonymisation of judicial staff in special circumstances (e.g. related to personal safety), but also establishing: *"(…) The identity data of magistrates and clerks may not be reused with the object or effect of evaluating, analysing, comparing or predicting their actual or supposed professional practices (…)."*[45] The recent Belgian act literally copied this provision, adding also 'lawyers' to 'magistrates and clerks'.[46]
As was pointed out by (de Vries), the French law does not differentiate in the goals of the processing of these personal data, while it might be reasonable to distinguish between the processing e.g. by law firms to optimize their litigation strategies and for academic or journalistic research.
In the Netherlands there is no specific legal provision regarding the processing of personal data of judges, so the general rules of the GDPR apply. An interesting case study can be found at *openrechtspraak.nl*, a website run by the Open State Foundation (OSF), a private non-profit organisation advocating an open government. OSF uses the open data from the decision database and extracts all the names of the deciding judges, which is a type of processing of personal data that can hardly find a base in Art. 6(1) GDPR.
As a next step OSF collects information from the judicial *name register*. This register is intended to improve the transparency and scrutability of the judiciary and has a firm legal base.[47] It contains data about the judicial career and side jobs of all judges. The name register can be found on a specific subdomain,[48] and has its own robots.txt,[49] disallowing all crawlers to index the register, except for the homepage. This robots.txt can be interpreted as the materialisation of the decision of the controller that making the contents of the register available as open data is incompatible with the goal the data have been originally collected and processed for (Art. 2(1)(g) of the current Act on the reuse of public sector information juncto Art. 6(4) GDPR). This in itself makes the reuse of the data already unlawful; when the AI-ODD enters into force, reuse of the name register will also be directly prohibited in the newly proposed Art. 2(1)(h) of the Act on reuse.[50] Regular search engines respect this robots.txt, but so doesn't OSF, by scraping all information from the name register.

---

[42] https://uitspraken.rechtspraak.nl/robots.txt, accessed on 26-11-2023.
[43] *The Judge Statistical Data Ban – My Story – Michaël Benesty*, on Artificial Lawyer, accessed on 18-12-2023.
[44] *France Bans Judge Analytics, 5 Years In Prison For Rule Breakers*, on Artificial Lawyer, accessed on 18-12-2023.
[45] Article L10 of the code de justice administrative
[46] See footnote 26.
[47] Act of 21 May 2012, Stb. 2012, 220, amending the Act on the legal status of Judicial Officers (Wet rechtspositie rechterlijke ambtenaren).
[48] https://namenlijst.rechtspraak.nl.
[49] https://namenlijst.rechtspraak.nl/robots.txt.
[50] See also the explanatory memorandum, which explicitly mentions this register.

As a third data processing step, OSF combines the information from both sources to expose them via a search interface on openrechtspraak.nl. Similar to the behaviour of Global24h.com in the cited Canadian case, the full contents of openrechtspraak.nl are exposed to Google and other search engines. This exposure results from not using a robots.txt file while offering a sitemap structure[51] that is clearly designed to facilitate web crawlers. As a result, when googled (intentionally, or by namesake) judges appear in the Google result list with a hit on openrechtspraak.nl. Finally, OSF also offers an API[52] for those wanting to reuse these personal data.

8 Final Remarks

Advanced technological possibilities have enabled court decisions to be published in growing volumes. In the Netherlands a programme is unfolding to enlarge the number of published decisions substantially, defended by the judiciary as a response to a broad societal demand (Naves et al.). This alleged demand though is mainly expressed by a small but vocalist group of transparency evangelicals, open data enthusiasts and big or legal tech companies. A recent report (Gisborne et al.) in the UK – a country that just recently showed initiatives for publicly disseminating court decisions on a larger scale – examined the public support for court data reuse. The result is worrying: many are afraid that sensitive information about the intimate details of their personal lives, made available to a court of law, by themselves or supplied by others, will be out in the open, uncontrollable and forever, for reasons unclear for them. Whatever the outcome of the much-needed – broad – societal debate, it should be conducted vigorously. Basically, this discussion touches on the question whether informational self-determination (Hornung and Schnabel) should be more at the core of the data protection regime, instead of the free flow of information and the obligations of controllers and processors. And it is also about the Rule of Law, since: *"The failure of the legal system to maintain the ancient balance between access and privacy will lead to the greatest danger of all – inhibiting citizens from participating in the public judicial system."* (Winn)

Since 1999, The Netherlands has been in the forefront of disseminating court decisions, with a rather broad publication policy – also for lower courts –, APIs for unlimited reuse, exposure to global search engines and an internal database unnerving privacy-aware members of staff. Basically without a legislative framework and without an adequate policy that has taken technological developments and evolving privacy legislation (especially the GDPR) into account. And I have demonstrated, when it comes to the GDPR and the public availability of court decisions, it is not just about the volume and not just about publishing them on a website. It is also about the quality of pseudonymisation, and also about all those other types of processing that are overlooked too easily but that need a proper legal base in legislative measures, with detailed safeguards and effective supervision. Prohibiting specific types of processing is also an option, of course.

Contrary to the situation in the Netherlands, public access to court decisions in Belgium has been quite poor in the past two decades (Van der Haegen). Legislation for improving the situation has been withdrawn twice. In the currently adopted act though, with very detailed input by the Belgian DPA,[53] a strict data protection framework is one of the cornerstones of the new regime.
It should be studied meticulously by their neighbours in the North.

---

[51] https://openrechtspraak.nl/sitemap.xml.
[52] https://openrechtspraak.nl/api_docs.
[53] DPA Belgium: Opinion on a draft proposal for an Act on the establishment of the Central Register for the decisions of the judiciary and concerning the publication of decisions, 13 May 2022. Text in French and Dutch.

Bibliography

Cadiet, L. *L'open Data Des Décisions De Justice*. Mission d'étude et de préfiguration sur l'ouverture au public des décisions de justice, Paris 2017. https://www.justice.gouv.fr/sites/default/files/migrations/portail/publication/open_data_rapport.pdf

Csányi, Gergely Márk, et al. "Challenges and Open Problems of Legal Document Anonymization." *Symmetry* 2021, 13, 1490. DOI: 10.3390/sym13081490.

Custers, Bart, et al. "Quis Custodiet Ipsos Custodes? Data Protection in the Judiciary in EU and EEA Member States." *International Data Privacy Law,* 2022, Vol. 12, No. 2: p 93-112.

de Vries, G.M. "Rechtswetenschappelijk onderzoek naar rechters door kwantitatieve analyse van jurisprudentie: waar ligt de ethische grens bij het profileren van rechters?" *RM THEMIS* 2021-2, p. 50-68.

Deuber, Dominic , Michael Keuchen, and Nicolas Christin. "Assessing Anonymity Techniques Employed in German Court Decisions: A De-Anonymization Experiment." *32nd Usenix Security Symposium*. 2023. https://www.chaac.tf.fau.eu/2022/11/10/usenix-security-symposium-2023/

Gisborne, Jennifer, et al. *Justice Data Matters; Building a Public Mandate for Court Data Use,* 2022. https://perma.cc/39PU-UVFV.

Groos, Daniel, and Evert-Ben van Veen. "Anonymised Data and the Rule of Law." *EDPL* 2020-4, p 1-11.

Hornung, Gerrit, and Christoph Schnabel. "Data Protection in Germany I: The Population Census Decision and the Right to Informational Self-Determination." *Computer Law & Security Report,* Volume 25, Issue 1, 2009, p/ 84-88. http://dx.doi.org/10.1016/j.clsr.2008.11.002.

Naves, H., et al. "Meer én verantwoord publiceren van gerechtelijke uitspraken." Nederlands Juristenblad, 2021/3258.

Spindler, Gerald, and Philipp Schmechel. "Personal Data and Encryption in the European General Data Protection Regulation." 7 (2016) JIPITEC 163 para 1.

van der Haegen, M. "Falende rechstpraakontsluiting is democratisch deficit." De Juristenkrant, 2016, 334, p. 15.

van der Sloot, Bart, Saskia van Schendel, and César Augusto Fontanillo López. *The Influence of (Technical) Developments on the Concept of Personal Data in Relation to the GDPR*. Tilburg: TILT, 2022.

van Opijnen, M. 2014. "Op en in het web. Hoe de toegankelijkheid van rechterlijke uitspraken kan worden verbeterd." Dissertation. Amsterdam UvA, 2014. Print.

Van Opijnen, M. 2021. *Alle uitspraken online? Hoe dan? Noodzakelijke ingrediënten voor een wettelijke regeling*. Ars Aequi 2021, feb., p. 127-135.

van Opijnen, M, and A Ivantchev. "Implementation of ECLI - State of Play." *Legal Knowledge and Information Systems - JURIX 2015: The Twenty-Eighth Annual Conference*. Ed. Rotolo, A. Amsterdam: IOS Press, 2015.

van Opijnen, M., et al. *On-Line Publication of Court Decisions in the EU. Report of the Policy Group of the Project 'Building on the European Case Law Identifier'* 2017. https://ssrn.com/abstract=3088495.

Vokinger, Kerstin Noëlle, and Urs Jakob Mühlematter. "Re-Identifikation Von Gerichtsurteilen Durch «Linkage» von Daten(banken). Eine empirische Analyse anhand von Bundesgerichtsbeschwerden gegen (Preisfestsetzungs-)Verfügungen von Arzneimitteln." *Jusletter* 2. September 2019.

Winn, Peter A. "Online Court Records: Balancing Judicial Accountability and Privacy in an Age of Electronic Information." *Washington Law Review* 79 (2004): 307-30.