Op en in het web: Hoe de toegankelijkheid van rechterlijke uitspraken kan worden verbeterd

van Opijnen, M.

[Link to publication](#)

Citation for published version (APA):
van Opijnen, M. (2014). Op en in het web: Hoe de toegankelijkheid van rechterlijke uitspraken kan worden verbeterd

SUMMARY

ON AND IN THE WEB

How the Accessibility of Judicial Decisions Can Be Improved

Due to the emergence of the internet, the dissemination of case law has undergone a radical change. Until late 1990s, commercial publishers had a monopoly on collecting, selecting, publishing and enriching case law. But the internet has made it possible for the judiciary to take the publishing of the case law into its own hand, both for internal use as well as to cater to the needs of the legal practice, scholarship and society at large. The dissemination of judicial decisions to the public serves both as a legitimizing function as well as the law-making function of jurisprudence.

However, the new opportunities evoke numerous questions: do all judgments have to be published, or does it suffice to publish a selection of them? And what should be the criteria for the selection? What is the extent to which judgments have to be rendered anonymous? Does it suffice to publish the bare documents, or should additional accessibility features be included so that the user is able to conduct a search effectively and efficiently in rapidly growing repositories?

After chapter 1 which is an introduction, in chapter 2 we discuss the legal framework relevant for the publication of case law on the internet. A distinction is made between three different concepts: the public pronouncement of the judicial decisions as such, the provision of case law and the accessibility of the published information. Data protection is discussed separately.

The public pronouncement relates to the legal status of the judicial decision. Judicial decisions must be pronounced publicly in order to prevent secret judgments and to enable the scrutinization of the judiciary. However, the daily practice – sanctioned by the highest national and European courts – is far less absolute and unconditional as seems to be prescribed by the relevant provisions of the European Convention on human rights and fundamental freedoms and the Dutch Constitution. Accordingly, a decision can be read aloud in open court, in full, or summarized, but a written judgment can also suffice, provided it is sent to the parties and available for those who can show a legitimate interest.

The provision of case law has had a passive character for a long time: a (specific) decision could be obtained from the court. On payment of a court fee, the ruling (generally) was provided, sometimes in anonymized or compressed form. This passive provision (i.e., one had to ask for it) is regulated in detail in various (procedural) laws, but the applicability of these rules to the publication of judicial decisions on the internet is questionable. Active provision exists in two forms: specific active provision and general active provision. The specific forms of active provision are entirely regulated by law. Examples are the active provision of judicial decisions to protect third parties (e.g. in cases of insolvency) or because of their law-developing nature. Publication of a judgment can also be ordered by way of additional punitive measure.

In the Netherlands, there is no formal legal framework on the general provision of judgments. The clearest instruction on publishing (a selection of) interesting cases is found in the ministerial memorandum 'Towards accessibility of public documents' from 1997. The most important instruction norms for the selection and presentation of case law can be found in Recommendation R(95)11 of the Committee of Ministers of the Council of Europe.

These, and other national and international sources, indicate the existence of an obligation to publish case law, albeit on a limited basis: a representative selection is sufficient. In the Netherlands, the selection criteria for publication are established by the judiciary itself. In 2012, the rather vaguely formulated guidelines originally drafted in 1999 were replaced by a more elaborate scheme in which different publication regimes apply to different (departments of) courts, procedures and case types.

Of the three concepts – public pronouncement, provision and accessibility – the latter is the most open. But it is not clearly defined anywhere. Moreover, the relevant legal instruments are of a secondary nature and generally do not specifically address access to judicial decisions. Yet some clear trends can be drawn from the available legal sources: case law should be made available for free, there shall not be any technical or legal barriers to reuse, they have to be identified in a unique and persistent manner, there shall not be any doubt concerning the origin of the published documents, the position of a decision within the entire court procedure must be visible, judgments have to be published timely and shall not be removed without good reason. Measures should also be taken to cope with the abundance of information: case law should be presented intelligibly and in a broader context, be available via a single website, be provided with metadata and be accessible and understandable in a cross-border context. Finally, a user must be able to find the relevant information in a large collection of case law as quickly and easily as possible.

The anonymization of judgments is governed by the Data Protection Act and the related legal framework, in particular an opinion of the Registration Board from 1997. This opinion states that all personal data that are not relevant for answering the legal questions or the interpretation of the law, must be removed from the judgments. The judiciary itself drafted detailed guidelines, not only prescribing which data have to rendered anonymous, but also stating that the anonymization should be done in such a way that the readability of the judgments is impaired as little as possible.

In an international comparative study we outline major differences in the way various countries publish judicial decisions, both with regard to the legal framework, the volume, the selection criteria used, the way in which publication and selection are organized and the ease with which the user can consult the case law published. However, the protection of personal data develops as a hard constraint on the publication of case law universally.

In comparison, it should be noted that in the Netherlands relatively much case law is published and the accessibility aspects score reasonably well.

Finally, in this chapter the trends and opinions in legal doctrine are discussed. Much of the legal literature on the general provision of court decisions over the internet dates back to the years when the technical possibilities were on a developmental stage. Facing unprecedented optimism about the possibilities of 'openness' and research, there was concern about the protection of personal data and the risks of information overload; depending on the author, the emphasis is either on the first or the second aspect. Some authors take an activist attitude, possibly losing sight of the borderlines between the themes of public pronouncement, provision, accessibility and data protection.

In chapter 3, we focus on the publication practice as it has developed in the Netherlands. First, we describe the recent history, in which two projects played a crucial role: firstly, the web portal Rechtspraak.nl with its case law database, and secondly Porta Iuris, an internal web portal of the judiciary, offering access to both internal case law collections and third party repositories. The organizational embedding and the information architecture of these facilities are outlined.

Next, the publication of case law on Rechtspraak.nl is tested against the legal framework described in chapter 2. To this end, first the policy as formulated by the responsible authorities is examined. Next, the question is posed whether the selection of case law as published on Rechtspraak.nl is sufficiently 'representative' – a concept difficult to objectify, and in this context actually confusing. We do this by comparing this selection with both the collection published by commercial publishers, as well as with internal statistics on the handling of cases. It is concluded that the published selection is sufficiently representative, although there are shortcomings with regard to some specific chambers. Moreover, the Netherlands lags behind on publishing relevant judgments of the European Court of Human Rights.

On the accessibility aspects, the score is quite well, especially since the completion of the project Nova Porta Iuris in June 2013. Since then availability of data for reuse has improved, formal relationships between decisions are made visible and judgments are identified by the European Case Law Identifier (ECLI).

At the end of this chapter, other dispute resolution institutions are discussed, such as disciplinary tribunals – which can publish their rulings since 2009 on Tuchtrecht.nl – the Dutch Foundation for Consumer Complaints Boards and the Netherlands Institute for Human Rights.

In the chapters 4 to 7, some specific topics related to the accessibility of case law databases are discussed in more detail. Chapter 4 focuses on the identification and citation of case law. First a theoretical framework is formulated. The first part of this framework is based on the Functional Requirements for Bibliographic Records (FRBR) of the International Federation of Library Organizations and Institutions (IFLA). The FRBR distinguishes four ontological levels. The first level is the 'work', an abstract level identifying the intellectual creation. The second level is the 'expression', also an abstract level that involves the (artistic or intellectual) realization of a work (e.g. a translation). The third level is the 'manifestation', which embodies the expression in a particular format. Finally, the fourth level, the 'item', is the single exemplar of a manifestation. The second part of the theoretical foundation describes an assessment framework with which identification systems for case law should comply.

Subsequently, both the Dutch and some foreign identification systems are assessed against this framework. The method commonly used to identify judicial decisions by means of 'parallel citations' faces numerous objections: it is not an identifier at the work level but at the expression level and it is also vendor-specific. The use of 'triples' (the combination of court name, date of judgment and case number) also faces numerous problems. The National Case Law Identifier (LJN) was an identifier at the work level, and offered important benefits,

especially in conjunction with the LJN-index, which was available on Rechtspraak.nl and also contained parallel citations.

In citation guidelines and daily practice, the LJN has gained ground rapidly. Opaqueness, finiteness, a non-full coverage and the absence of cross-border interoperability gave rise to the development of the ECLI-system. The design, European embedding and the implementation of ECLI are described in some detail.

At the end of this chapter, the identification of legislation and other legal sources is discussed.

Chapter 5 deals with the problematic issue of selection, which already existed when case law was published in paper magazines only. Different (proposed) selection criteria are discussed, as well as the selection problems faced in daily practice. It appears that the testing of individual decisions against the selection criteria is often more difficult than formulating the criteria itself. Since more decisions are actually published, being selected is becoming less of an indicator for legal relevance. So, due to the massive publication of case law, 'selection' is changing from an input problem to an output problem, and should therefore be treated as a relevance issue from the science of information retrieval.

In a theoretical treatise on the concept of 'relevance', we distinguish between five 'manifestations' of relevance. The first is system relevance, which describes the extent to which a particular document matches a query in purely mathematical terms. Secondly, there is topical relevance, which expresses whether a document pertains to a specific subject matter. With cognitive relevance, we determine whether a document is also perceived as relevant by the user. Situational relevance expresses how a document contributes to the solution of the problem which prompted the user to start his search initially. Finally, there is domain relevance, referring to the importance attached to an information object by the whole community (the 'crowd'), regardless of the context of a specific user query or the subject of the document.

For the purpose of the research in chapter 7, in chapter 5 some methods to measure relevance are also discussed, with an emphasis on social network analysis.

Chapter 6 deals with the 'contextuality' of a judicial decision. Accessibility of judgments can be improved if they can be viewed in their legal context – e.g., their position on a broader line of jurisprudence – or if they can be searched for by the paragraphs of legislation which they quote. Partly as a result of the myriad different ways that are used by judges and academic authors to quote case law and legislation, this contextuality is hard to achieve through the use of out-of-the-box functionality of search engines. Because the sheer volumes make manual tagging impossible, specific software is needed to establish the links between judicial decisions, (national and European) legislation and legal literature.

These software mechanisms are described in this chapter. To detect case law citations, first of all those text strings are recognized that may be part of a citation: names of courts, dates, (possible) case numbers, parallel citations and LJN's. These 'citation strings' are then normalized and finally canonicalized, a process in which the validness and uniqueness of

case law references are established. Consequently, a citation network is built, with which a decision can be viewed in its legal context. The time required to find where in legal doctrine or case law a particular judicial decision is cited, can thus be reduced from hours to seconds.

To detect references to legislation a slightly different methodology is used. First, a database was built containing all known official titles, abbreviations and spellings of local, national, European and international laws and regulations. All these strings are subsequently recognized in the judgment text, after which a special parser converts those citations to a canonical form and detects the specific paragraphs that are referenced. For European legislation, which is often cited with document numbers, a specific solution was devised, and local aliases – the habit to use, e.g., 'the statute' or 'the convention' in subsequent citations of an initially fully quoted regulation – are recognized correctly. Using this technology, collecting all case law citing, e.g., a particular paragraph of a European directive, is significantly accelerated and improved.

Finally, this chapter describes the content and development of the research database, which was not only used for the work described in this chapter, but also for the research in chapter 7. This database contains over 850.000 case law documents and more than 550.000 files of legal literature.

In chapter 7, we develop a model with which the domain relevance of judgments can be expressed on a five-point scale. In this 'Model for Automated Rating of Case law' (MARC) the public life of a court ruling is divided into three phases. The first one is the publication phase, the short period after the pronouncement in which the judgment is published for the first time (on Rechtspraak.nl) and it is made available for scrutiny by the 'legal crowd'. In the 'transition phase' there are the selection by law journals, the writing of annotations, the insertion of the decision in 'continuous literature' (such as commentary editions, online manuals and wikis) and the first citations of the judgment in case law and 'one-off literature' (such as articles in law reviews and dissertations). Finally, in the 'citation phase' the judgment is 'settled', it is only cited in case law and one-off literature.

For these three phases different models are developed. For the citation phase, an algorithm with eight independent variables ('predictors') was devised: the age of the decision, the hierarchical position of the court that issued the ruling, the field of law and the five 'crowd variables': publication and annotation in periodicals (where both the number of magazines/annotations as well as the standing of those magazines are taken into account), citation in continuous literature, citation in case law and citation in one-off literature. For the latter two a weighted moving average is being used. As the dependent variable (the 'regressor'), we use the odds that the judgment will be quoted in one-off literature and case law within the next three years.

In the publication phase, the crowd variables cannot be used, because the crowd has not been able to study and assess the decision on its general legal importance yet. Therefore, for this phase mainly endogenous variables are used, i.e., data contained in the judgment itself, such as the number of judges deciding the case, the number of outgoing references to

legislation and case law and the length of the decision. In addition, the field of law and the hierarchical position of the court are also used. Moreover, the publication of the decision on Rechtspraak.nl is considered as a predictor; whether or not an accompanying press release has been issued on this website turned out to be statistically irrelevant.

For the transition phase – for which a period of 100 days is considered appropriate – a weighted average of the scores in the publication phase and the citation phase are used.

Because the numerical results of the algorithms – calculated with the generalized (non) linear model – are hard to interpret for end-users, these were reduced to five ascending classes: MARC-1 for trivial judgments, MARC-5 for landmark cases. When testing the same set of decisions in both publication phase and citation phase, the endogenous variables turn out to be a good predictor for the opinion of the legal crowd during the citation phase. It is also established that the endogenous variables have no added relevance in the citation phase. Apparently these factors are (implicitly) taken into account by the legal crowd.

Some final considerations can be found in chapter 8. In addition to a summary of the findings and an answer to the research questions, the importance of the type of research which was conducted in the chapters 6 and 7 for legal practice and scholarship is discussed. To put this into perspective, we take into account the problematic relationship between the natural sciences on the hand and legal practice and scholarship on the other, but also the appealing prospects of the legal semantic web. The rapidly increasing availability of digital legal source materials (both quantitatively and qualitatively) offers entirely new opportunities for legal knowledge management and – interdisciplinary – legal research.