The Expressiveness of Legal Information from an Information Security Perspective

Selene M. Morrow

Abstract

In technical and legal contexts, the information security concept is commonly discussed as containing a number of elements, such as confidentiality, integrity and accessibility. However, these discussions often neglect the need to preserve the semantic content of information, failing to address this need through the inclusion of an appropriately representative information security element.

It is the purpose of the present work to elaborate on an additional component of information security termed expressiveness. Usage of this term, or at least variations thereof, is not altogether uncommon in a number of fields – particularly in the context of information standards. However, this usage is often characterised by a high degree of technical specificity, which leads to narrow focus and an inability to properly relate expressiveness to other elements of information security. Additionally, there exists no identifiable consensus among researchers, authorities and industries when it comes to defining what expressiveness really entails.

This article aims to provide a conceptually sound expressiveness concept, which is particularly suitable in the context of information security issues in the legal field. This is done through several steps:

• Traditional concepts of information security are correlated to various processes of legal information such as creation, retrieval, reproduction and transformation.

• It is shown that the process of information transformation is directly related to a change in the semantic content of information (semantic discrepancies).

• It is argued that such semantic discrepancies may be present even in the absence of other apparent threats to information security (such as compromised integrity). Furthermore, the potential threats of semantic discrepancies in legal information are elaborated upon.

• The term expressiveness is presented as an indicator as to the absence or presence of semantic discrepancies in legal information.
1 The Traditional Sub-Elements of Information Security

It is often said that legal information security encompasses a number of sub-elements. Most commonly, these are described as confidentiality, integrity and accessibility.

Confidentiality refers to the need to restrict access to certain information, so that only authorized parties have access to it. Which parties are authorized may be defined in law (such as statutes and precedents concerning personal data and secrecy). It may also be determined in other contexts, e.g. by way of internal company guidelines or contractual agreements.¹

The integrity of data refers to factors such as accuracy, completeness and currency.² It can be threatened by intentional and accidental events, such as corruption, alteration and deletion. Computer viruses, data intrusion (i.e. “hacking”) and other such phenomena constitute typical threats to information integrity.

Accessibility also constitutes a central factor of information security.³ Information is obviously of little value for individuals, companies and authorities if access to the information is restricted. Such restrictions may take the form of limited opening hours or substandard networks, which hamper the

¹ In our digital era, confidentiality is often ensured through technical measures such as encryption. However, it should be noted that encryption, in its basic form, is quite an ancient phenomenon – going back at least as far as the Roman times. The Caesar cipher and many other encryption techniques throughout the ages are discussed at great length in Kahn, The Code Breakers: The Comprehensive History of Secret Communication from Ancient Times to the Internet, Scribner, New York, 1997.

² The integrity of data can be said to be equivalent to “data quality” in many ways. However, the term “integrity”, by virtue of its linguistic qualities, seems to encompass a greater breadth of factors than does the term “data quality”. Specifically, the former term is more suitable for describing situations where there is no data to speak of, and thus no possibilities of performing a quality evaluation – such as in situations of incompleteness due to deletion of data.

³ A linguistic difference can be noted here between availability and accessibility of information. The term “availability” can be used for the purpose of denoting that certain information does or does not exist. Accessibility, on the other hand, refers specifically to whether access to this information is possible in a specific context. A lack of availability logically infers a lack of accessibility. The opposite is not true however, since access to information may be hampered even in cases where the information is available in some form (e.g. to a select number of individuals with specific security clearance). While more general in character, the term “availability” is less suitable than “accessibility” when it comes to discussions in the field of information security. A lack of information availability can stem from numerous different factors (e.g. political and economical), many of which are not directly related to information security factors. From an information security perspective, it seems more worthwhile to focus on existing information and the manner in which its access is restricted through technical and administrative measures. Building upon what has been said above, the term “availability” is often utilised in political goal-statements. See e.g. Online Availability of Public Services: How is Europe Progressing? – Web Based Survey on Electronic Public Services, Report of the Sixth Measurement, Capgemini, 2006.
ability of users to reach an authority or company during certain times. Other times the accessibility restrictions are related to confidentiality concerns, making it necessary to strike a balance between the need to preserve confidentiality and the need to promote accessibility (e.g. disputes concerning access to public records which include sensitive personal data).

In a European context, the importance of accessibility is a recurrent theme in e-government projects. E-government refers to the utilisation of IT-tools for the purpose of increasing the efficiency, speed and quality of governmental functions. This can be exemplified by possibilities of registering businesses and filing taxes online.

Increasing accessibility is a significant goal of legal automation – the partial or complete performance of legal functions through computerised means. Obviously, automation may diminish reliance on staff, which likely translates into increased ability to provide information quickly and independently of typical concerns (such as the hours of the day and geographical distance).

Returning to our analysis of the components of legal information security, one can mention the additional elements of accountability and non-repudiation. These two elements overlap considerably, as the existence of one largely infers the existence of the other. Furthermore, it should be noted that the two terms “accountability” and “non-repudiation” concern possibilities of denying responsibility for certain actions (e.g. rescinding a contract by claiming not to

---


5 The notion of highly automated authorities is closely related to the concept of decision support systems (DSSs). These are discussed further below (2.3 and 2.4).

6 Nonetheless, the two terms are not completely synonymous. It is possible that an individual is unable to repudiate a certain action, but can still not be held (completely or partially) accountable for it. This can be exemplified by a situation where it is clear that a contract has been signed by someone underage or otherwise incapable of being subject to a contractual obligation. Thus, non-repudiation is a necessary, but not sufficient, precondition for accountability. On the other hand, an element of accountability must generally be said to infer an element of non-repudiation. Holding an individual accountable for an action would seem to require that he/she cannot repudiate this action – at least if we assume that this determination of accountability is made on the basis of fairness and accuracy.
be the originator of an electronic signature). One could argue that the broader term "identifiability" would be more suitable in this context, as this would better emphasize an ability to take responsibility for actions in the face of claims against the contrary. Such matters, however, are better discussed elsewhere, as they concern terminological issues rather than the practical ones underlying the present article. No matter which term one chooses to describe accountability, non-repudiation and/or identifiability, the three concepts remain facets of the same general function.

Having looked at the traditional elements of information security above, we will now turn to the primary focus of this article. This concerns an additional element of information security which is often neglected, if not always in practice then at least when it comes to categorisations of information security components in legal contexts. In order to discuss this elusive additional element, we need to take a closer look at what it means to create, retrieve, reproduce and transform legal information – actions which can be subsumed under the heading of “information stages”. An analysis of these four stages obviously requires that the term “legal information” is itself defined. However, as we will see, establishing such a definition is far from straightforward.

2 Legal Information

2.1 Legal Information Creation and the Authority Criterion

The creation of legal information refers to situations where someone (be it an individual or some sort of private or public body) produces a piece of legal information. Generally, the individual or body in question must have some sort of authority to produce such information, or the information cannot be considered "legal" in character. This can be referred to as an authority criterion governing the dividing line between "legal information" and "other information".

While the above definition may, at first glance, seem reasonably uncomplicated, one quickly encounters difficulties in the establishing of strict boundaries when it comes to which kind of authority is necessary for

---

7 Electronic signatures utilise encryption technology for the purpose of achieving digital representations of an individual's identity, which are comparable in legal value to handwritten signatures. In Europe, they are regulated through Directive 1999/93/EC of the European Parliament and of the Council of 13 December 1999 on a Community Framework for Electronic Signatures. In this context, it can also be noted that a technical definition of non-repudiation and accountability may not always be in line with a legal interpretation of said concepts. The electronic signature directive aims to address this overlap by incorporating technical specifications and requirements directly into legal text. That notwithstanding, there still exists a large divide between legal and technical professionals when it comes to these issues.

8 The same can also be said for the common term “traceability”, which seems to emphasise accountability aspects. An individual must generally be “traced” only if they are either unintentionally lost, or intentionally hiding from view. As has previously been noted in relation to the terms non-repudiation and accountability, the term traceability is ill-suited when it comes to describing possibilities of self-identification (as one cannot generally be said to trace oneself).
something to be considered legal information. These terminological complexities cannot be easily (if at all) remedied, as the term "legal information" has not itself been given a precise and universally accepted definition. Nonetheless, it is possible to demarcate specific situations where the creation of legal information must generally be considered to have taken place. The most obvious examples are the creation of legislation through parliamentary action, as well as the creation of precedents through the workings of courts. In both these cases, it is hardly debatable that legal information is in fact brought into existence - excluding these manners of information from the umbrella of "legal" would obviously make the entire term "legal information" meaningless.

Other situations are far more difficult to define as to their role in the process of legal information creation. Legal doctrine clearly exemplifies these definitional problems. For instance, is it possible to laymen to create legal information? As an example, one can consider a journalist writing an article about a certain legal phenomenon. Is such information not truly "legal" in character? What if the journalist is also a lawyer, or used to be a lawyer but switched careers? A follow-up question would be whether or not it is necessary to possess legal education in order to be defined as a potential creator of legal information. Assuming the latter case, what level of legal education is then necessary? It is obvious that no clear answers to these questions can be attained. Thus, any usage of the (admittedly highly practical) concept of “legal information” must be tempered by an awareness of the uncertainties inherent in this concept.

2.2 Legal Information Retrieval and the Dividing Line Between Authoritative and Factual Information

The authoritative legal information discussed above is obviously necessary to retrieve in some way before it can be utilised for purposes of legal decision-making, drafting of contracts, representation and advising of clients, etc. It is hardly a surprise to anyone that paper-based services of legal information provision are quickly losing significance in relation to legal databases – particularly those with an online presence. Such legal databases can be found both in the private and the public sectors. Generally it can be said that state-
maintained initiatives tend to provide a barebones approach to their legal information retrieval services, presenting the original sources with few to no improvements and modifications. For obvious reasons, commercial databases in the private sector tend to provide considerably more functionality, often going beyond mere retrieval of authoritative legal information and on to assisting users when it comes to understanding, interpreting and applying the located material. In this context, it is important to observe the commonly thin line between information retrieval and information transformation, a topic which will be discussed further below (2.4).

Above, the uncertainty inherent in the "legal information" concept has been discussed in the context of legal information creation. Legal information retrieval brings to light another definitional issue related to the legal information concept itself – namely whether legal information can attain its legal character by virtue of de facto being retrieved for legal purposes. Consider, for instance, personal data concerning a defendant in a court case. In order to evaluate suitable criminal sanctions in this case, it may be necessary to ascertain the defendant's age, criminal background and other legally relevant factors. A similar situation concerns retrieval of personal information through an online service utilising so-called intelligent forms. Examples of information that can be thus collected include an individual’s name, age and income. Intelligent forms generally constitute interfaces between the users of a computer system and another party such as an authority or a company. While their level of sophistication can vary, such forms generally allow for the automatic collection of inputted information for use in subsequent processes (such as legal decision-making). At the very least, this can include sorting the information in different database fields, by virtue of associating these fields with the fields in the form. On a higher level of complexity, the information in intelligent forms can be put to use for the purpose of reaching partly or completely automated decisions, e.g. concerning eligibility for certain tax benefits. One can argue that factual information of this kind, due to its...

Potential issues related to transformation of legal information are outlined below (2.4).


11 The term “system” is not easily defined in an exhaustive manner. A classic definition of the term has been formulated in Bertalanffy, General Systems Theory: Foundations, Development, Applications, Penguin Books, Harmondsworth, 1968. While Bertalanffy applied this theory in the field of biology, its key points are also relevant to most any fields of knowledge, such as those of law and IT. Specifically, system theory focuses on the interdependence of elements in a given context, according to the notion that “the whole is greater than the sum of its parts”. In the present article, we will utilise the term “system” in the context of functions that are in some way the result of combinations between software (e.g. applications) and hardware (i.e. physical components of computers).

12 It should be noted that the handling of personal data is restricted by legislation in most jurisdictions. In a European context, one should note Directive 1995/46/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.
relevance to legal information such as statutes and precedents, must itself be
defined as a kind of “legal information”. This terminology also seems the most
suitable, as any other approach would seem to neglect the decisive importance
personal data and other such factual information can have on legal processes.

Additionally, attempting to establish a dividing line between "legal
information" on the one hand, and "factual information" on the other, would
cause excessive definitional problems when it comes to information which is
simultaneously factual and quasi-legislative. Examples of this situation include
employment contracts in the context of labor law disputes between employers
and employees. As another example, one can mention contracts between
consumers and companies and their role in civil litigation processes. In both
these instances, the contracts constitute factual elements (i.e. the facts of the
case), while simultaneously containing rules which must be interpreted in a
manner normally associated with legal statutes and other legal information.
Furthermore, conflicts between such contracts and relevant
legislation/precedents may bring to light situations where all these types of
information have to be reconciled (e.g. by taking into account whether some
aspects of the legislation are dispositive, and the contract thus must be
considered valid in these regards). The close ties that often exist between
authoritative and factual information clearly illustrate the problem in reserving
the term “legal information” strictly for information of the authoritative
category. Consequently, it is in the best interest of consistent terminology to
define legally relevant factual information as a kind of legal information in and
of itself.

Nonetheless, for purposes of clarity, one can draw a line, albeit an inevitably
rough one, between legal information which has achieved its legal character by
virtue of the related creation process itself (as discussed above in 2.1) and
information which instead constitutes general facts, which derive their legal
status from their relation to other legal information in a given context (such as
a process of legal decision-making pertaining to an individual). We will call
the first type authoritative legal information, whereas the latter type will
simply be referred to as factual legal information. It is true that such a
definitional approach somewhat reintroduces the problem of drawing sharp
boundaries between the two types of information, as has recently been
discussed. On the other hand, it enables us to include both information types
under the wider umbrella of "legal information", thus achieving a reasonable
terminological compromise.

2.3 Legal Information Reproduction

The reproduction of legal information refers to situations where the source of
the information may already be in the possession of the individual or body
performing the copy. One example is the common process in which public
authorities and private companies produce digital versions of paper-based
legal information to which they already have access.
It may often seem a straightforward matter to reproduce legal information. Before the computer era, such reproduction generally involved either a handwritten or typewritten reproduction of the material in question and, particularly as time progressed, some manner of photographic reproduction of pages (e.g. through a photocopier). From the perspective of information security, reproduction of this kind mainly gave rise to information integrity concerns. Particularly in the case of handwritten and typewritten copies of material, there is obviously a risk that the individual performing these processes will unwittingly omit or alter parts of the original information. These risks are diminished by processes of photographic reproduction, even though one can still point to integrity concerns by way of partial or complete omissions of pages (e.g. faded borders leading to unreadable footnotes) and other such problems. Technological evolution has since given way to new methods by which to store and represent legal information, such as databases and even decision support systems (DSSs) which handle part of the information interpretation process. These methods will be discussed further below, in relation to analysis of the process of information transformation (2.4).

First, however, it is important to consider the line between legal information reproduction and legal information retrieval. Taking this line into account is important – particularly from a terminological point of view. Consider, for instance, an individual (let us assume a legal professional of some kind) goes down to the local library, locates a seemingly relevant book on legal matters and proceeds to copy a number of its pages. The individual then brings the pages to his/her workplace and utilises them in the context of work tasks of a legal character. Does this constitute information reproduction or information retrieval? Before answering this question, let us increase the level of involved technology somewhat, and instead consider an employee who accesses databases for the purpose of locating legally relevant information. It is hardly disputable that such database access must be considered information retrieval. However, does it make any difference to our definitions if we instead assume that the person copies the database information over to a document on his/her computer hard drive (e.g. an ongoing word processor document)? The latter example obviously conforms to an everyday use of the term “reproduction”.

13 The terms “store” and “represent” are used here to denote examples of reproduction methods. While the two terms are closely related, they are not synonymous. Representation refers to the manner in which information is reproduced onto a medium, but where – due to the character of the medium – the information must be reshaped in such a way that one cannot speak of a literal reproduction. As mentioned above, typical examples of representation can be seen in DSSs, which cannot generally handle legal information (e.g. statutes, precedents and doctrine) in “raw” form, but relies on such information being incorporated into the system, for instance through programming code.

This sheds light on the fact that we are dealing with a two-step process in these situations, including two distinct but closely interrelated elements – the preceding retrieval and the subsequent reproduction. Thus it would appear that the change of medium must be seen as the decisive factor for determining when legal information reproduction has indeed taken place.15

2.4 Legal Information Transformation

The above definition of information reproduction as a change of medium helps us envision yet another line which must be drawn between concepts – one between reproduction and transformation. In order to better understand the transformation concept, we need to focus on the technological methods by which reproduction can take place. We have briefly touched upon some of them above, and there it was observed that some of the more recent methods – such as representation of legal information in databases and the development of DSSs – merit special attention. What, then, are the distinguishing features of these methods of information reproduction, which set them apart from more traditional methods, such as typewriting and photocopying? The answer to this question relates to the degree in which different forms of information reproduction simultaneously lead to the information undergoing some degree of transformation.

In order to analyse these issues further, it is important to determine exactly what is meant by transformation in this context. Consider a legal database, which contains full-text versions of court case summaries. Let us further assume that each case summary is accompanied by references to legal doctrine, which can help shed light on the matters underlying the case in question. These references are recommendations, which are not present in the original case summary but only in the representation of this summary in this particular database. An individual accessing the case report will thus be provided with guidance as to a suitable manner in which to interpret certain facets of the case, but the guidance stems from a different source than the case itself. For instance, it is possible that some team of legal professionals was involved in the development of the database, their knowledge and/or personal experiences constituting the basis of included recommendations. In yet other situations, it is possible that the development of the database, including recommendations as to relevant doctrine, has been undertaken by other types of professionals, such as journalists. Regardless of how that may be, the common bond between all these examples is that something has been added to a source, which was not present in the original “barebones” representation stemming from the relevant authoritative source. Thus, the manner in which the cases have been incorporated into the database includes not only an element of reproduction, but also an element of transformation.

15 When an individual accesses information in a database, this information is, in a way, copied into his/her mind – stored in memory for future use. Nonetheless, it seems terminologically unsuitable to define this process as including an element of information reproduction, until the information is de facto transferred to a new physical medium accessible to others. In other words, the mind, the brain, personal recollection and such factors are not defined as mediums in this context.
Of course, it could be argued that the information itself is not changed in these situations, but that further information is merely added to it. Nonetheless, it still seems suitable to define such additions as a transformation of the entire body of relevant information (e.g. the entirety of a case report), since the additions are bound to colour the perception of the information as a whole. What is meant by this is that even if the first part of a database entry contains an exact quote of the original case report, reproduced verbatim from the official court versions, the additions present in the second part result in a material change also when it comes to the first part of the entry. For instance, the references to “recommended reading material” found in the second part of the entry may invite a certain perception of the original case report, whereas a different selection of materials might have resulted in the user of the database forming a differing view. Obviously, to provide an extreme example, an additional sentence in a database entry saying “the above information is only valid until January 2007” will obviously result in a substantial transformation of a database user’s perception concerning the database entry as a whole. Characterising the sentence in question as a mere addition would be misleading, even if it may be perfectly apparent to any database user that the sentence does not stem from the original authoritative source but was introduced by the system developer.

What has been said above means that one could even define handwritten additions to paper documents as substantial transformations of legal information – to say nothing of the numerous methods of reshaping information that the digital era has brought about. In this context, it is relevant to note the potential consequences of legal information constantly being conveyed in new and original ways. This can often overshadow the fact that the information is not merely being reproduced in a “refined” manner – in the sense that it is made more accessible and easier to understand – but that it may be simultaneously transformed in a way which affects related perceptions and interpretations.

A clear example of such transformation of information can be seen in different forms of automated systems in the legal field. This refers to systems which not only offer information retrieval capabilities but in fact suggests possible interpretations of the material and may even perform part of the interpretative process for the user. This is obviously closely interrelated to the database additions discussed above, but takes the guidance process a step further by not merely suggesting recommended material but even going as far as de facto applying it to an individual situation. As mentioned earlier, systems with the purpose of aiding users in the process of reaching some kind of decision based on legal material are often referred to as legal decision support systems (or DSSs), and constitute clear examples of a strong overlap between information reproduction and information transformation.
We will now elaborate on the concerns surrounding the transformation of legal information by way of discussing an element of information security which we can term *expressiveness*.

## 3 Analysing the Concept of Expressiveness

### 3.1 The Expressiveness Concept in the Field of Information Standards

When information security is envisioned, analysed and discussed in various situations, it is unusual that the concept of expressiveness is truly taken into account when it comes to its crucial role as an additional information security element. Nonetheless, it should be observed that the concept is fairly common in other fields, particularly that of information standards.

A short presentation of information standards is in order, before their relation to the expressiveness concept can be outlined. This will take its starting point in a discussion of XML (Extensible Markup Language). To put it simply, XML is a type of language used to describe information standards. Information standards are, in turn, a set of standardized code-tags which provide information segments with additional specifications. These can concern layout, e.g. “<bold>” and “<title>”, as is illustrated by the Hypertext Markup Language (HTML) so commonly utilised in the creation of web pages. However, information standards can be used for many other purposes, such as denoting the structure and content of documents. The possibilities are, in a way, unlimited. Information standards can point out the location of specific legally relevant facts, such as “<WitnessName>” and “<MonetarySum>”. They can even specify the semantic qualities of texts such as “<HighRelevance>” and “<MitigatingCircumstance>”.

16 The scope of the present work does not allow for an extensive discussion of other fields where the expressiveness concept (or variations thereof) is utilised. However, it should at least be noted that the concept is often referenced within the field of logic programming. See e.g. Dantsin et al., *Complexity and Expressive Power of Logic Programming*, in Computer Surveys (CSUR), Volume 33, Issue 3, pp. 374 - 425, ACM, New York, 2001.

17 In 2000, HTML was made an international standard (ISO/IEC 15445:2000).

18 For extensive analysis of these uses of information standards, see Magnusson Sjöberg, *Critical Factors in Legal Document Management: A study of standardised markup languages*, Jure, Stockholm, 1998.

19 Of course, this does not imply that information standards alone can actually bring about the manner of functionality which they express. Specifically, some interpreter – generally a form of information system or program – must be able to interpret the tags, and act upon the encapsulated text accordingly. The most obvious example of such an interpreting program is the web browser, which is able to “understand” and process information in HTML tags. When it comes to information standards with a higher degree of semantic specificity, it generally becomes necessary to rely on systems with some capability of automated reasoning (e.g. DSSs) for the purpose of interpreting the tagged information.

20 It must be kept in mind that information standards can only be efficient if their use is, indeed, standardised. This may seem like a self-evident matter, but an identical use of information standards “across the board” in a certain context can often be difficult to bring about. When it comes to state authorities, inefficiency and lack of communication can often lead to a simultaneous development of distinct information standards which are not compatible. Private
When information standards are utilised in order to denote their structure and content of texts, *expressiveness* is often discussed as an important related factor. Specifically, the expressiveness concept is often utilised for the purpose of denoting the descriptive power of information standards. This expressiveness is directly related to the DTDs (Document Type Definitions) and Schemas. To put it simply, DTDs are specifications which contain the elements and attributes allowed in a XML document. Schemas are optional ways by which to formulate such specifications, which possess certain advantages over DTDs. For instance, XML Schemas utilise the XML language itself, whereas DTDs use a specialised language. Another advantage, more relevant to our purposes, is that XML schemas are generally said to be *more* expressive than DTDs, and the expressiveness of XML schemas has often been made the object of specific study. It should be noted that many of these studies apply a narrow definition of the expressiveness concept. The one thing common to most all discussions concerning expressiveness in the field of information standards is that they refer to the ability of such standards when it comes to describing as many matters as possible. Somewhat ironically, however, the term “expressiveness” is not used in a clear, standardised manner in the field of information standards, but can refer to a multitude of issues surrounding the possibilities inherent in such

companies, on the other hand, often have their own reasons (generally related to competitiveness) for not favouring dissemination and general awareness of their utilised standards.


22 It may not be altogether correct to define this as a universally acclaimed advantage, as the resulting syntax of XML schemas has been described as “intricate” and unreadable even to XML experts – precisely due to the fact that it is based upon XML. See Liu, Mengshi, *DTD schema: a simple but powerful XML schema language*, in International Journal of Web Information Systems, Volume 4, Issue 4, pp. 465 - 483, Emerald Group Publishing Limited, 2008. Here, aforementioned criticisms are presented, and a recent XML schema language by the name of DTD Schema is suggested as a way to reap the benefits of both DTD and XML schemas while avoiding their weaknesses. It is specifically pointed out that DTD schema is much more expressive than XML schema when it comes to describing XML elements and attributes.

23 See e.g. Martens, et al., *Expressiveness and Complexity of XML Schema*, in ACM Transactions on Database Systems (TODS), Volume 31, Issue 3, 770 – 813, New York, 2006. Here it is also expressly pointed out that “XML Schema directly addresses most of the shortcomings of DTDs, and, in particular, is more expressive than DTDs” (p. 771, my italics).

24 As an example, reference can be made to a document prepared for the W3C XML Schema Working Group, shortly before the WC3 issued its approval of XML schema in 2001. This document discusses the “expressive power” of XML schema. However, the term “expressive power” is used strictly from the point of view of formal language theory. Specifically, it is used for the purpose of denoting the extent to which features of DTDs (and other former technologies) can be recreated through XML schemas, as well as the extent to which XML schemas offer descriptive powers which lie beyond the capabilities of DTDs. The W3C document in question offers the following concise definition: “Given two formalisms P and W, P is more powerful than W if any language which can be defined using W can also be defined using P, and there is at least one language which can be defined using P which cannot be defined using W. Another way to say this is to say that the set of languages whose rules are expressible by P is a proper superset of the set of languages whose rules are expressible by W.” See Replicating DTD Functionality Using XML Schema, W3C XML Schema Working Group, 2000 (http://www.w3.org/2000/04/20-xmlschema-dtdissues.html at 1.1).
standards. One conclusion that can nonetheless be drawn from the use of the expressiveness concept in the field of information standards is that it refers to the accuracy which a certain information representation method allows.\textsuperscript{25} We will keep this conclusion in mind as we move on to discuss the concept of expressiveness in a legal framework.

3.2 Evaluating Expressiveness in the Context of Authoritative Legal Information

We will now focus further on the meaning of expressiveness when it comes to authoritative legal information and the manner in which it may be represented in legal databases, DSSs, etc. In order to do this, we must consider the relation between the expressiveness and the integrity of legal information. In other words, can a line be drawn between the integrity of legal information being compromised, and the information’s expressiveness being endangered? We will answer this question by looking at situations involving threats to integrity and expressiveness, from the perspective of the various information stages discussed above (see chapter 2). By contrasting these two information security elements, the significance and independent role of the expressiveness concept will be clarified.

Above (2.1) it has been noted that legal information creation, as the concept is defined for our purposes, refers to processes during which authoritative legal information comes into being. Obvious examples include parliamentary legislative proceedings and the decision-making actions of courts. Threats against information integrity at this stage include situations of oversight and misconduct during the performance of such processes. For instance, a legal act may not be properly documented in a way which represents what was in fact decided at parliament when the act was passed.\textsuperscript{26}

When it comes to the element of expressiveness and its relation to the stage of legal information creation, it must be kept in mind that expressiveness refers to the accuracy with which certain information is represented. This is consistent with the aims of the present article, as well as with the common use of the term in other fields according to what has been discussed above (3.1). If we define authoritative legal information as the underlying material which subsequent representations are based upon, it thus becomes apparent that authoritative legal information cannot itself be subject to threats against expressiveness. In other words, authoritative legal information is used as the

\textsuperscript{25} The concept of representation has been discussed briefly previously in this text (see supra note 13).

\textsuperscript{26} It must be emphasized that we are here dealing only with inaccuracies stemming from the original process which brought a certain act, precedent or other authoritative legal information into being through some form of official documentation. Inaccuracies introduced at a later stage by third parties would not be associated with the stage of information creation, but with the stages of information reproduction and/or information transformation. This is consistent with our definition of the information creation stage as strictly concerning the creation of authoritative legal information (see 2.1 above).
reference point by which other representations of this information (e.g. privately owned and maintained databases) are judged as to their expressiveness. To evaluate the expressiveness of authoritative legal information in the perspective of utilised terminology would thus be meaningless.

Moving on to the legal information retrieval stage, it must be noted that new technologies in our digital era have led to a considerable increase in the number of integrity threats. While it is not the primary purpose of the present work to elaborate on such threats, computer viruses, illegal access to systems and data interference can be mentioned as prime examples which are not only given considerable media coverage, but have also been made the object of international legislation.\(^{27}\) On the other hand, the concept of expressiveness is wholly related to the manner in which certain legal information is represented by an authorised party (e.g. a team of system developers). Thus, it is not concerned with breaches of integrity by unauthorised third parties.

Reproduction of legal information can give rise to integrity concerns of the same kind discussed above in relation to legal information retrieval.\(^ {28}\) Expressiveness cannot be affected by the reproduction of legal information – as long as the reproduction in fact remains a reproduction and does not include aspects of transformation. In this context, it must be observed that even seemingly minor alterations in the presentation of certain material can induce changes in the way this material is perceived by individuals.\(^ {29}\)

The three information processes discussed above – information creation, information retrieval and information reproduction – can all be subject to integrity concerns. However, they are not possible to evaluate in terms of expressiveness. The transformation of legal information differs from previously discussed processes in that it can involve threats to both integrity and expressiveness. The integrity concerns are closely related to those discussed in other contexts, but with an added dimension. It is a well-known fact in the world of computerised handling of data that the more times information is reworked, redesigned and otherwise manipulated, the greater...

---

\(^ {27}\) An important piece of legislation to observe in this context is the Convention on Cybercrime (ETS No. 185), drawn up by the Council of Europe, which entered into force in 2004.

\(^ {28}\) Obviously, integrity concerns may not be the only legal considerations involved in the reproduction of legal information. While generally not relevant when discussing information in the form of statutes and precedents, copyright factors may be highly significant when it comes to legal doctrine and its reproduction in databases, applications and other contexts.

\(^ {29}\) Some studies have touched upon the manner in which the presentation of legal texts affects their understanding. See e.g. Gunnarsson, Legtexters begriplighet : en språkfunktionell studie av mediestämmelagen (Eng: The Comprehensibility of Law-Texts: a Pragmatic Study of the Act on the Joint Regulation of Working Life), Liber, Lund, 1982. Findings in related fields also support the theory that presentational factors can affect perceptions quite significantly. See e.g. Zuckerman and Chaiken, A heuristic-systematic processing analysis of the effectiveness of product warning labels in Psychology & Marketing, Vol. 15(7), pp. 621-642, New York University, 1998. Among other layout issues, this study analyses the difference in perception resulting from texts of different colours.
the potential for integrity issues. As legal information is transformed in some way, it is often subjected to considerable changes in structure and content, and mistakes can thus often creep into the resulting representation (commonly program code).

However, there are also situations where no apparent syntactical error exists, but legal information is nonetheless transformed in a way that leads to a change in semantic character. This can be referred as semantic discrepancies between the underlying piece of authoritative legal information and the manner in which it has been represented. Utilising the terminology presented in this article, we can define such situations as threats to the expressiveness of the representation.

The need to uphold expressiveness as an element separate from integrity can be best exemplified by instances of legal information transformation where little, if anything, in the original source material is de facto altered, but some kind of functionality is merely appended. For instance, consider a DSS which provides guidance to case handlers at a public authority when it comes to deciding applications in the field of social benefits. Let us furthermore suppose that, for a given set of circumstances, the system contains 40 precedents that could potentially be used as basis when the system decides the application. Our next assumption, which may not altogether inaccurate in the context of decision-making at public authorities, is that case handlers statistically rely primarily upon the precedents which appear in the first page of results that the system provides. However, who is to decide the order in which the results are presented?

Here we have a feature which cannot possibly be traced back to the underlying authoritative source of legal information – at least not until we live in an era where parliament passes program code, rather than acts and other traditional forms of legislation.\(^30\) The relevance ranking of the search engine is instead directly associated with the manner in which the system was developed, rather than with any constitutionally approved legislative and decision-making processes. It is possible that lawyers and other legal professionals may be employed at the system development stage, for the purpose of providing expert opinions concerning suitable ways in which to rank the precedents in the database. More commonly, the relevance ranking functions may be directly related to certain algorithms, e.g. concerning the frequency with which certain terms appear in the retrieved documents. While the former approach may seem more desirable than the latter in the perspective of legal accuracy, neither approach is derived from an authoritative legal source. Thus, they both give rise to situations where the expressiveness of the information in the database is

---

\(^30\) Many researchers claim that such a time is not too far away, and that computer code has de facto replaced law as primary regulatory force in many facets of today’s society. This view has been perhaps most famously presented by researcher Lawrence Lessig. See Lessig, *Code and Other Laws of Cyberspace, Version 2.0*, Basic Books, 2006.
compromised – it does not accurately express the content of the underlying authoritative legal information.

How is this dilemma to be remedied? It is obviously necessary to provide some method for ranking the results, and it would seem the methods discussed above would in any case be preferable to a totally random order of presentation. This may well be true, which is why it is often inevitable to sacrifice the faithfulness to the character of the original text – i.e. the expressiveness of the legal representation – to some extent, particularly when dealing with computerised tools offering additional functionality that cannot possibly be found in the underlying material which they represent. Nonetheless, what is important to observe is that this situation, and others like it, can potentially lead to information security threats. These threats concern the manner in which authoritative legal information is transferred from state bodies (e.g. parliaments and courts) to individuals (either in their private or their professional capacity).

To further illustrate the point, we can expand upon the example above, and assume that the DSS in question is employed by every public authority in an entire nation, for purposes of providing guidance within a specific legal field. Consequently, the relevance-ranking mechanisms inherent in the DSS would, in essence, achieve quasi-legal status despite not being passed by a democratic body. This shows how the manner in which authoritative legal information is represented can constitute an obvious information security concern, even when the integrity of the information has not been compromised. Consequently, the need to uphold expressiveness as an information security element distinct from information integrity becomes particularly clear.

Despite what has been said above, it should be noted that the line between threats to integrity and threats to expressiveness may often be challenging to draw. It is possible to rely on the above distinction between syntactical errors and semantic errors to a certain extent, but it would not be possible to uphold a strict division between these two categories. For instance, we can envision a DSS with the purpose of assisting judges by providing suggestions for appropriate sentencing. Let us further assume that the DSS relies on a flawed representation of the underlying legal act, as it does not take into account certain mitigating circumstances found in this act. In attempting to define whether the integrity or the expressiveness of the original act has been compromised, we are faced with a dilemma. We could define the situation as a breach of integrity, by referring to the fact that some part of the act was completely overlooked when the system was developed, or possibly that the information was somehow erased or corrupted and thus led to the system behaving inadequately. On the other hand, it is also possible to define the issue as one of expressiveness, by reference to the fact that the act was interpreted in a way contrary to its meaning, thus altering its semantic qualities. If the act in question is challenging to interpret, it may be difficult, if not impossible, to distinguish between these two scenarios. While it may not be necessary to
establish such a distinction for practical purposes, it is nonetheless important to observe the difficulties in doing so.

3.3 Evaluating Expressiveness in the Context of Factual Legal Information

Above (2.2), we have discussed how information retrieval can often concern factual information, such as personal data. There it was also said that this type of information could be subsumed under the general concept of “legal information”, by virtue of its relevance to a legal process.

Our definition of the expressiveness concept so far has only concerned representations of authoritative sources, e.g. risks pertaining to the transformation of such sources into programming code. However, seeing as legal information can also be of the factual kind, it is necessary to extend the expressiveness concept accordingly.

When evaluating the expressiveness of factual legal information, real world facts must obviously function as the starting point against which any representations are judged. For instance, imagine a field in an intelligent form, which has the purpose of collecting the full name of an individual, separating the name into given name and surname and subsequently including this information in a database. If the form is flawed in its construction, or even in its instructions to users, there is a risk that the collected information becomes incorrect (e.g. through given name and surname being mixed up in the database). In this case, the representation in the database is inaccurate when compared to the original facts. However, from an information security perspective, it cannot be said that this constitutes a breach of integrity, as the information has not been corrupted. The inaccuracies were introduced in the system through a communications failure between the user and the system, not through any subsequent information security threat. Thus, it seems terminologically consistent to define the threats surrounding such accuracies as expressiveness threats related to the representation of factual information – in this particular case pertaining to the manner in which the factual information is translated into database entries.

As another example, on a higher level of technical specificity, we can consider a DSS with the purpose of evaluating an individual’s eligibility for social benefits. Such systems often function through posing a series of multiple choice questions to users, which are then correlated and cross-referenced with different sections in legal sources. The goal of the questions is to reach a point where a more or less exact answer to a user’s query can be provided. However, such multiple choice questions must obviously be designed well enough, as to avoid a case of miscommunication between the user and the system. For instance, the question “is the individual currently suffering from a disability” is not always easy to answer with a simple “yes” or “no”. Such an answer may often require that the user is completely familiar with the legal definition of “disability”, since most cases are far from black and white. DSSs may often attempt to counteract definitional problems of this kind by providing users
with additional information which is necessary and/or useful when answering the questions. Such information can e.g. be introduced in a “popup window” or a separate frame in the user interface. However, it is obviously important to strike a good balance between accuracy and user-friendliness. It goes without saying that systems cannot rely extensively on pure text quotes from legal sources for the purpose of assisting users in answering DSS-type questions. This can ultimately negate the very purpose of the system, as it would reflect any interpretative difficulties back to the user. Thus, we can see that the need to uphold the expressiveness of factual information may pose considerable challenges in the context of legal system and database development.

4 Summary and Closing Words

As we have seen above, the use of DSSs, sophisticated databases and systems of legal information retrieval can often, either intentionally or unintentionally, bring about instances of legal information transformation. In general, the more extensive the functionality offered by the database or the system, the greater the level of transformation involved. Certain systems even perform part of the interpretative process for users, thus considerably altering the manner in which legal work has traditionally been carried out and de facto reshaping professional roles and redefining legal requirements and responsibilities.

The development outlined above brings about a need to expand upon traditional information security notions. The information security concept has traditionally been divided into categories such as integrity, confidentiality and accessibility. These categories seem to focus only on the information stages which have conventionally been in the foreground – information creation, information reproduction and information retrieval. Taking into account the ever-increasing importance of legal information transformation, it appears necessary to extend the taxonomy of legal information security with a concept that is specifically geared towards the risks involved in such transformation. This article has elaborated upon a concept of expressiveness for purposes of identifying, describing and evaluating information security risks related to the transformation of legal information into computerised representations utilised by systems and databases.

In many ways, it is clear that transformation of legal information is the way of the future. Many researchers in the field of law and information technology even foresee a not too distant development where code will in fact replace law as the primary regulatory mechanism in society. It could in fact be argued that, in many regards, this is already the case today. Nonetheless, it is quite possible that the strength of this trend will increase exponentially over the course of the next few years. As has been hinted at above, this development may possibly even result in coding decisions commonly becoming political in nature. Parliaments may start to forego the traditional legislative process altogether in instances when enacting finished computer code is considered a more straightforward and effective measure.
For these reasons, it is necessary for developers of legal information systems to understand the great need to involve legally trained professionals in system development processes. In a sense, legal computer system development can be seen as embodying a culture clash, where communication across professional boundaries becomes paramount to efficient co-operation and high quality results. Through this article, I have aimed to clarify the need for such communication, by raising awareness of the risks associated with flawed legal transformation. An understanding of the role of expressiveness in the context of legal information security constitutes an important step in addressing such risks.

5 Sources

Doctrine


**Legislation and Reports**


**Websites**


Westlaw – at http://www.westlaw.com