Cecilia Magnusson Sjöberg*
The Melting Pot Paradox of Structured Documents

1 A platform for true action

Sticking to documents as a basis for a discussion on development trends in information society might be regarded as somewhat old-fashioned. But for all the growing impact of multimedia on the legal domain, the document concept still plays an important role as a basis for legal discussions, above all because law is still produced in the form of text entities commonly referred to as documents. Typical document instances are acts and ordinances, decided cases, contracts and clauses, conventions and so forth. Actually, a more appropriate expression in this context would be virtually tangible documents, meaning that these kinds of objects representing law may occur in electronic as well as in paper formats. The introduction and application of IT has, furthermore, led to a state of art where the boundaries of legal documents are not necessarily defined beforehand. The Swedish Principle of Publicity, for instance, is characterised by a right of access to dynamic constellations of data.

This presentation sets out to show that structured documents accomplished by means of standardised markup languages ought to be regarded as a highway route on the map of an information security characterised by security. The somewhat obscure title “The melting pot paradox of structured documents” is chosen in order to capture the embedded and intended contractions in the approach advocated here, viz order achieved by encumbering text with <tags> (mainly elements and attributes) and static views of structures as a basis for dynamic actions.

The grooves and moves in this context are trusted public and business activities in terms of information retrieval, knowledge management, automated decision-making, e-commerce, etc. Trust implies information security conventionally comprising the criteria of availability, confidentiality (secrecy), integrity, accountability and non-repudiation. Clearly, these building bricks of security have legal implications both in terms of system design and management as well as regards applicable rules and regulations. In the following it will be sufficient to let these concepts serve as a general framework for the discourse.

---

* *Cecilia Magnusson Sjöberg* is professor at the Faculty of Law, Stockholm University and Royal Swedish Academy of Sciences Research Fellow. She gained her doctorate in 1992 with a thesis on legal automation in the Swedish public sector. She is currently managing a research project that investigates the possibilities of cross-fertilisation of advanced methods for security enhancement and applications of XML in the legal domain.

2 The mission

Structured documents, as a tool for legal validity in a security context is the mission. This statement may in itself be regarded as a content-heavy expression that can be the object of an analysis resulting in marked-up elements *and* attributes. A supporter of standardised markup languages is thus expected to take a dynamic approach to text, while those not yet informed of the potentials of the W3C Recommendation XML (Extensible Markup Language, http://www.w3.org/XML/) might still be stuck in thinking only about flat representations of key words attached to the beginning of a document or search words inserted in an (inverted) index file.

3 Why markup languages?

3.1 XML explained

It is high time now for a brief explanation of what document markup is all about. An XML document may be *well formed* or governed by a DTD (Document Type Definition) or schema. A *document type definition* defines the composition of a set of documents (e.g. laws and court cases). It contains information about document elements, the logical order of these elements and their frequency, etc. A DTD is expressed in XML and may be stored in a data file outside the document.

There are different ways of explaining the underlying meaning of a DTD. One could focus on the purpose of a DTD as a method of *structured information description in context*. This implies that a DTD does not necessarily have to be related to a certain type of document but rather to some particular kind of information. A skilfully designed DTD with corresponding markup makes it possible to adjust the use of a particular document to a variety of purposes, i.e. without later having to change the markup. XML then plays the role of an enabling tool, making it possible, for instance, to find information that is of interest on a specific occasion. This is an indication of how important a preparatory *document analysis* is.

A *schema* can be generally described as a specification or formal definition of the constraints on the content of an XML document, aiming at both structure and functionality. One way to specify a schema is to use a DTD, but XML schemas can model other kinds of structured data as well and are in principle more expressive. An important feature of an XML schema is the possibility of integrating database functionality and communication between applications. A major purpose of an XML schema is indeed to make it support data typing (integer, date, etc.) and thereby facilitate XML data interchange with conventional database systems. XML schemas are written in XML and have been developed for use on the Internet and are therefore co-ordinated with other W3C specifications.

A *document instance* is the encoded document itself containing data (e.g. legal text), *markup* (document element tags) and a DTD reference (if not present in the document). The markup elements surrounding the text are called *tags*. In the simple example below the tags are displayed in bold characters. The value of attribute type ID is shown in the ‘article tag’. ‘A3-95-46-EC’ here stands for Article 3 in the EC Data Protection Directive.

```
<ARTICLE ID='A3-95-46-EC'>
<ARTTITLE>Scope </ARTTITLE>
<ARTNO>Article 3 </ARTNO>
   <PARA> 1. This Directive shall apply to the processing of personal
data wholly or partly by automatic means, and to the processing otherwise
than by automatic means of personal data which form part of a filing
system or are intended to form part of a filing system.
>  </PARA>
...
</ARTICLE>
```

In comparison with HTML, the dramatic difference is that while HTML aims at presentation of text on a (computer) screen, a major purpose of XML is to allow for semantic expressiveness. Furthermore, the HTML DTD consists of a predefined tag set whereas an application based on XML is open to any kind of customised vocabulary. Of utmost importance is the inherent validation component of an XML application governed by a DTD or schema. In practice this means that a marked-up document is validated against the predefined logical constraints (such as decided order of elements) and the predefined number of occurrences of particular elements.

3.2 XML and security

So far so good, but is the XML approach secure and what makes it at all worth investigating in terms of legal implications? The purpose here is to convey a strategy for answering this question. The aim is thus not merely to satisfy curiosity but to make the approach serve more practical interests. A legally founded checklist of *XML related security-enhancing factors* would no doubt enhance the rule of law and thus promote trust in system design based on XML solutions (see further Section 5.3 below).

One starting point is that modern document management requires co-ordination in order to meet demands for efficient production, supply and use. Apart from general needs to improve recall and precision when retrieving information, there is also reason to consider, for instance, knowledge management attempts and exchange of business data in networks of various kinds.

XML has a potential to function as a lever and a sound basis for all of these developments in modern information society. In this context it should be mentioned that XML also has a profound impact on *substantive law* itself, in particular in the fields of contract law, intellectual property rights and privacy protection. For instance, XML-messaging quite often comprises personal data processing in a legal sense (see the EC Data Protection Directive, 95/46/EC). It concerns requirements of consent from data subjects to collect, store and disseminate personal data. Furthermore, modern e-business models make it necessary to consider information duties, e.g. that the identity of a service provider must be clarified according to the EC Directive on E-commerce (2000/31/EC). Liability issues are also relevant in terms of an analysis of who is responsible for damages emerging as a result of the abuse of a transferred authentication.

As indicated above, it all boils down to trust in global digital information and a need for *legal information security* in open as well as in

closed computer-based networks. Every organisation, be it a private enterprise or a public authority, needs to reflect upon the handling of documents governing internal as well as external actions. One highly important question, for example, is how far XML may support message authentication and electronic measures to prevent distortion of (document) content. The concept of authority here covers a wide variety of actions, e.g. authorisations to enter into contract, and law enforcement.

Bearing in mind the initially mentioned checklist approach, *the group of addressees* or "*who will benefit*" may be described in the following way. To begin with, secure use of XML is relevant for commercial actors as well as for representatives of the public sector. This can be instanced with *buyers* who, in a procurement situation, are dependent on clarification of legal conditions governing a particular situation. In a *vendor* perspective, the use of a checklist may be regarded as a business opportunity in terms of a legally founded security branding of offered solutions. Enhancing legal awareness among *politicians and public officials* for the purpose of efficiency, foreseeability, uniformity, openness, etc. is another obvious advantage.

3.3 XML as a tool

The expression "XML as a tool" connotes, to begin with, *markup languages in a broad sense,* not least including SGML (Standard Generalized Markup Language), considering the impact of this ISO standard, dating back to 1986, on applications which are still running today (ISO 8879:1986). A markup language may be utilised for representation of structures and contents, styling and communications. This implies that not only the core XML W3C Recommendation is of interest in this context, but also related standardisation initiatives such as XSL (Extensible Stylesheet language), SOAP (Simple Object Access Protocol), etc. Bearing this in mind, XML ought to be regarded as a symbol for a *system development approach* commonly including data management in terms of text.

At one stage of the relatively short historical development of this type of applied information technology, the SGML community was a pretty closed one, not particularly amenable to discussions, for instance, concerning the pros and cons of various database technologies. Today the situation has changed in that XML can be said to play a central role in more or less any technical solution involving web technologies, telecommunications as well as conventional electronic data processing and to some extent also techniques having their origin in artificial intelligence.

All of this may no doubt be elementary to the already experienced user of the above-mentioned family of standards. Practical experience has shown, however, that a common misunderstanding at the management level of an organisation, be it a private enterprise or a public agency, is that XML (in the broad sense) involves choosing a particular system design and possibly even software product. Representatives of the industry as well as other promoters of standardised markup languages thus have the educational task of explaining the underlying ideas of a non-proprietary approach to data management. Otherwise there is an obvious *risk* that such a lack of understanding may turn out to be a major *obstacle* for widespread use of XML. In fact there may be legal advantages associated with awareness of

the inherent capacities of XML. The choice of system development approach as such may have an impact on a court's assessment, for instance, of whether an organisation's archival system is to be regarded as accurate or negligent in terms of meeting legal requirements of evidence by keeping track of version-dependent legacy data. The pharmaceutical and vehicle industries are typical examples of branches heavily burdened by legal requirements of documentation. However, it may not be a trivial task in a litigation situation to explain to a court just how the use of XML manifests a party's legal awareness.

To summarise, although XML may be described as a tool, it is not just any kind of tool. It is not a physical object like a pen or a paper. Nor may XML be described as a mechanical mechanism resembling, for instance, the functions of a typewriter. XML is instead a tool with *strong infrastructural potentials* closely interlinked with IT-support for information retrieval, document management and knowledge management. From a legal point of view, this deserves particular attention.

4 Infrastructural changes

The term infrastructure is often used to describe the fundamental functions of society. It can refer to both 'hard infrastructures', such as the road system, or 'soft infrastructure', such as social systems and various types of information systems. The basic components of a legal infrastructure, which may be regarded as 'soft' according to the above-mentioned classification, include various forms of (a) data processing, (b) documentation, (c) communication, and (d) organisational forms.

The introduction of information technology into society has brought about dramatic changes to all these components. For example, *data processing*, which was a *manual* activity in the past, was transformed step by step during the 1970s and 1980s into automated data processing of cases. Today, automation of administrative activities, in the sense of legal decision-making based on wholly or partly *automated routines*, can be said to be a characteristic feature of administrative procedures. Another type of legal data processing takes place in connection with the development and conclusion of contracts. The technical possibilities of electronic conclusion of contracts with the whole world as a market place warrants a discussion in this area concerning the fundamental legal principles underlying offers, acceptance, evaluation of evidence, etc. *XML* obviously has a role to play here as a *tool for improved legal system management* considering its potentials for handling version-dependent text units over time. Earlier generations of lawyers would naturally associate the concept of '*documentation'* with physically demarcated paper documents, which could be geographically located. In the age of the Internet this view is not longer valid. It is no longer obvious that documentation consists of *paper documents*. In many cases it may come in different forms of *electronic documents*, which are carriers of declaratory acts, proprietary rights, criminal contents, etc. XML clearly mirrors this development. Mention should here be made of such initiatives as *XML Signatures* that explicitly address the need for incremental signatures, which, for instance, may be of relevance in a situation of successive drafting of contracts.

In a similar way (voice based) *analogical* communication services are used less frequently in legal work. Both civil servants' *communication* with the citizens, and lawyers' contacts with their clients are increasingly dependent instead on *digital* and mobile services. In Sweden, for instance, the comprehensive systems for the dissemination and collection of information by the authorities are based on a strategy that may be referred to as a kind of *XML-labelling*. The system for Dissemination and Collection (Sw. SHS) constitutes the public administration's investment in order to create a general communication link to secure information exchange through the open Internet. In contrast to electronic trading systems, which are usually designed with a focus on business transactions in a certain sector, SHS constitutes a general platform, which has not been especially programmed for a certain sphere of activity.

As regards *organisational forms* we have a strong tradition of working with nationally *well-demarcated larger and smaller entities*. This is especially clear as regards the information system of public administration which has developed in harmony with nationally defined government authorities which are divided into central and local organs, etc. Information technology as such and the Internet as a concept have provided leverage for loosening up boundaries between authorities as well as national demarcation lines. The private sector may be characterised by even more *all-embracing, network-based and global organisational forms* in recent years. It is evident that *XML supports or rather is an integrated part of this infrastructural shift*. Obviously this gives rise to a whole series of substantive law issues, for instance, how to apply privacy legislations to transborder flows of personal data.

5 Interaction of law and IT

5.1 Regulatory management

In the era of IT-supported document management there is a growing need for version control in a long-term perspective. Document markup, including linking techniques of different kinds, is attractive as a general value-adding method. At the same time, the introduction and widespread use of more and more advanced digital document management systems is resulting in a very complex environment for text handling. Furthermore, *open systems* are a major development trend in today's communications networks. One important concern, therefore, is how best to secure trails of authorisations, alterations included. More precisely, this is a matter of information security policies mirroring the norms that govern an organisation, such as who has a right of access to what, without knowing beforehand who will be claiming this right of availability.

Considering that the *major characteristics* of normative documents are complex, interdependent text units, shifting in content over time, interpretable only in context, we can extract one key issue, and that is the question of a *methodological approach* to regulatory management. If the challenge in terms of a required *infrastructure* is overcome, we can indeed expect the added value so often promised by vendors. There is otherwise an obvious risk of turnback or perhaps even failure.

The cornerstones in a *system development approach* meeting the *fundamental requirements of modern regulatory management* are (a) document markup, (b) information security, and (c) legal awareness. XML naturally represents the core method as regards *document markup*. From the point of view of regulatory management, XML offers vital possibilities of transparent modularity in a structural context. The conventional understanding of *information security* is that it comprises availability, confidentiality (secrecy), integrity, accountability and non-repudiation. Ongoing work at IETF (The Internet Engineering Task Force) on securing web-based documents will of course serve as an important input. In this context the focus of attention is on the XML Signature and Encryption initiatives. The application of a cryptographic method of progressive (incremental) security enhancement may serve as supplement. Finally, *legal awareness* is required in terms both of methodological aspects of legal system development and of substantive law issues related to the use of electronic signatures, electronic evidence, etc. The last mentioned perspective might also be expressed in terms of *possible legal validity as proof of actions* of different kinds.

5.2 Electronic signatures as an illustration

There is no doubt that *electronic signatures* and other means of secure electronic messaging are becoming established in society. The main question of law is whether, how, and to what extent electronic signatures can be given the same *legal effect* as handwritten signatures. This will ultimately lead to the question of the *evidentiary value* of electronic signatures. This presentation does not aim to provide a detailed analysis of these questions in the light of different jurisdictions. Here it will be sufficient to say that in general civil law has a relatively limited number of formal requirements concerning handwritten signatures. Typical examples where such signatures are required include consumer credits and real-estate purchase. In family law, testamentary dispositions and marital property, agreements are often not valid without handwritten signatures. In administrative law it is more common to require signing, since processing of cases often presupposes submission of signed documents.

The implementation of the EC Directive on a Community framework for electronic signatures (1999/93/EC), which had to be enforced by the Member States by 19 July 2001 at the latest, clarifies certain legal situations. The overall objective of this Directive may be said to be to co-ordinate the legal and technical work of the Member States as regards electronic signatures, removing in this way the obstacles to the internal market, especially as regards e-commerce. The EC Signature Directive contains provisions relating to the legal effects of electronic signatures and the organs that may come to be able to offer electronic certificates verifying the genuineness of such signatures.

However, there is still a lot of work to be done before computer transactions can be performed on a daily basis with the help of modern information and communications technology in a *sufficiently secure way*. The latter expression refers to the need of minimising uncertainties as regards both different legal issues and practical (partly technical) circumstances surrounding electronic handling of documents. It may also be

said that business models and administrative traditions have not yet been fully adapted to the modern, more secure methods of information exchange.

Notwithstanding the above, the Directive can be said to be innovative from the European perspective in that electronic signatures under certain circumstances have been granted legal enforceability (see Article 5), which was not the case earlier in those countries which lack the principle of free examination and evaluation of evidence.

Applicable *Nordic law* shows in this respect that when the legislator uses the term 'written', electronic communication may be allowed to take place. The requirement of written procedure is posited here primarily as opposed to oral procedure. When a statute contains expressions such as 'signature' or 'that must be signed by the party in question', or the like, then this cannot be taken to mean that electronic documents or signatures may be allowed to replace traditional paper documents and manual signatures at the present stage of technical development. However, special rules or established practice may permit the use of electronic form after all.

As regards *evidentiary value*, the state of the law is a bit clearer in the sense that there is no doubt that the principle of *free examination of evidence* characterises Nordic law. In general terms, this means that there are no limitations on the sources of evidence that may be used at a trial, and also that a judge is not bound by any special regulations regarding the way in which different types of evidence shall be evaluated. Nor does the legislator seem inclined to introduce any general *rules on the burden of proof* which would be dependent on the medium used in a given case. The important factors are instead considered to be the parties' internal relations, the character of the legal document, etc. One can therefore conclude that Nordic courts encounter no formal obstacles to considering *system evidence* in the sense of electronic documents, electronic signatures and other components of information systems.

5.3 XML related security-enhancing factors

The discussion above boils down to a need for a strategy to handle legal uncertainty characterising the information society of today here illustrated by the EC Signature Directive. This manifest need for a legal strategy may in practical terms be transformed to a focus on *XML related security enhancing factors*. The table below presents an overview – not an exhaustive list – of what is here referred to as *XML characteristics, means and legal incentives*.

1. XML characteristics	2. Means	3. Legal incentives
Non-proprietary format	Inherent in any XML application	*Public sector:* Accessibility in a long-term perspective *Private sector:* Legal requirements of keeping track of legacy data
Quality assurer in document production	Validated documents by means of DTD:s and schemas	*Public sector:* Public information supply (possible state responsibility) *Private sector:*

		Commercial products (avoiding liability)
Quality assurer in document distribution	One single repository as a basis for customising production on, e.g. CD-ROM, on-line, print	*Public sector:* See above *Private sector:* See above
Container of legal directives	Markup vocabularies	*Public and private sectors*: Availability differentiator (Official or secret data) - Authority indicator (Mirroring acting parties authorities) - Property rights administration (Labelling of rights and its owners)
Secure electronic messaging	For instance, XML signatures or mathematical approaches to incremental signatures of structures and marked-up documents	*Public and private sectors:* Normative as well as other legal requirements of authentication, validation and signed documents (text entities)

The approach described above serves as a basis for the so called SLIM Project – Secure Legal Information Management – carried out at the Swedish Law and Informatics Research Institute (IRI) at the Faculty of Law, Stockholm University. The project will run during the period 2002-2008.[1] The SLIM project is founded mainly on previous practical and theoretical experiences of using SGML in the legal domain – the Corpus Legis project together with expertise from the Department of Information Theory at the University of Lund and the Swedish Institute of Computer Science (SICS).[2]

Because of the early stage of commercial tools that combine XML capability and digital security enhancing techniques, one aim of the SLIM Project is to have an impact on the development of future commercial XML tools for legal purposes.

6 Concluding remarks

This presentation is based on the standpoint that XML may be regarded as *a tool for legal validity in a security context*. A point is made of the fact that

1 http://www.juridicum.su.se/slim.

2 http://www.juridicum.su.se/iri/corpus.

XML ought to be understood broadly and that the *tool* metaphor has implications beyond trivial physical and mechanical ones. The term "*docware*" may in this context be used as a label for XML-related technologies clarifying the chosen approach to document management in a given situation.

In terms of general development trends we have reflected upon how XML has become an integral part of *modern infrastructures* with obvious legal implications. More precisely, this has a bearing on modern means of data processing, documentation, communications and organisational forms.

The fact that there are still so many legal uncertainties in terms of lack of foreseeability concerning legal validity of actions of various kinds calls for special attention. A pragmatic approach is presented in terms of regarding various *XML characteristics as security enhancing factors*. The attraction of combining XML with conventional security-enhancing methods lies in the need for transparent, content-dependent and context-sensitive management of legally relevant text units over time. Legal awareness in this context may indeed enhance trust in the information society.