Tools for Collaboration Between Transnational NGOs: Multilingual, Legislative Drafting

McIver Jr., William J.

National Research Council Canada
Conseil national de recherches Canada

Institute for Information Technology
Institut de technologie de l'information

NRC·CNRC

TOOLS FOR COLLABORATION BETWEEN TRANSNATIONAL NGOS: MULTILINGUAL, LEGISLATIVE DRAFTING *

W. McIver, Jr.
August 2004

* submission for the *International Colloquium Communication and Democracy: Technology and Citizen Engagement*, August 4 - 6, 2004 Fredericton, New Brunswick, Canada. NRC 48066.

Canada

Tools for Collaboration Between Transnational NGOs: Multilingual, Legislative Drafting

William J. McIver, Jr.

National Research Council Canada
Institute for Information Technology
e-Government/e-Citizen Group
46 Dineen Drive
Fredericton, NB E3B 9W4 Canada

e-mail: Bill.McIver@nrc-cnrc.gc.ca

Submission for the *International Colloquium Communication and Democracy: Technology and Citizen Engagement*, August 4 - 6, 2004 Fredericton, New Brunswick, Canada

Abstract

Non-governmental organizations (NGOs) face a broad spectrum of barriers to effective transnational cooperation (Ó Siochrú 2003). One critical barrier is the lack of ready access to software tools that facilitate transnational, multi-lingual, collaborative work.

As an example, Civil Society's drafting processes for the World Summit on the Information Society (WSIS) have been very complicated, tedious, and prone to error. Inputs are received from many NGOs or caucuses composed of NGOs. These inputs may consist of commentary or specific recommendations for language in some consensus document. All such inputs must be reconciled for inconsistencies, debated, and placed into a structure for the overall document.

Complicating the process further is the fact that NGO communities are now often distribute across multiple languages. There are, for example, six official languages in WSIS. Truly democratic debate over document revisions is severely hampered until translations of a draft have been produced.

The goal of strengthening transnational networks within civil society must, therefore, include the development of ICT tool sets that solve both the technical and social problems involved in managing formal or semi-formal democratic processes. A number of content management systems now exist that might be extended and adapted for this purpose, but no fully functional system as such exists.

A critical factor in providing this type of tool is the use of a free software model.[1] Proprietary solutions to various aspects of the problems of collaboration and multi-lingual drafting exist, but they are expensive for most NGOs to acquire and operate.

This paper will present the context in which advanced collaboration tools for NGOs is needed. It will discuss general system requirements and provide technical background. Some current technologies that provide parts of the functionality needed to support greater collaboration among NGOs will be discussed. The paper will conclude by presenting a high-level technical architecture for such a system.

[1] The term "free software" is taken in this paper to follow the definition of the Free Software Foundation (http://www.fsf.org), which is a more sound intellectual property regime for making software available to NGOs than the ambiguous term "open source."

This paper is in part a reflection on selected technical issues related to the Social Science Research Council Report prepared by Seán Ó Siochrú: *GLOBAL GOVERNANCE OF INFORMATION AND COMMUNICATION TECHNOLOGIES: IMPLICATIONS FOR TRANSNATIONAL CIVIL SOCIETY NETWORKING*. We are concerned here with the development of software tools to improve transnational, multi-lingual, collaborative work among non-governmental organizations (NGOs). This paper recognizes the spectrum of barriers identified in the Ó Siochrú report faced by NGOs (3). These include barriers to access and maintenance of solutions based on information and communication technologies (ICTs), as well as barriers between ICT-enabled NGOs and those that are not. This paper chooses to focus, however, on select issues of collaboration related to among NGOs characterized in the report as relatively well-endowed and ICT-enabled, which are either themselves transnational or participate with other NGOs in a international settings. We are interested, in particular, in issues that reside in the top two layers described in the Ó Siochrú report:

- Layer 1 -- Physical Access and Enabling Tools and Resources; and
- Layer 2 -- Generating, Retrieving and Using Content.

The Context

It is recognized that ICTs have played a role in the significant increase in the number transnational NGOs in the past two decades (Ó Siochrú, 2003:1). In the context of networked computing, this has included electronic mail (e-mail), bulletin board systems, the World Wide Web, interactive communications facilities such as Internet Relay Chat (IRC) and Instant Messenger from yahoo.com. Since the advent of the Web, a new class of systems has also

emerged -- content management systems (CMS) -- which make the authoring and publishing Web content more efficient. These technologies have enabled greater collaboration among organizations, but they alone do little to directly support and monitor collaborative processes. Such processes must still be managed to a large extent by the human "agents" within the transnational networks that are created by and between NGOs. Human agents should, of course, not be subject to management by machines, but arguably many collaborative processes can be improved through intelligent assistance by ICTs, particularly as transnational networking becomes more complex. ICTs already exist and are evolving. They are built upon the technologies in the physical access and content generation layers described in the Ó Siochrú report (7-23).

Greater Complexity in Transnational NGO Networking & Intelligent Assistance for Collaboration

Klein (2003) and Miller (2003) have both recently studied the increasing role of NGOs in summits sponsored by the United Nations and other organizations over the past three decades. There has arisen the so-called "multi-stakeholder" model for participation, which is ostensibly designed to involve civil society and other sectors in the inter-governmental-directed processes involved in summits. The preparatory processes involved in these formal settings have raised the complexity of NGO collaboration to a much higher degree. This is due to the increasing numbers of NGOs, greater geographic and linguistic diversity between NGOs, and the corresponding complexity that these increases bring within these formal frameworks. For example, NGO networks that wish to participate at this level may have to work against ever more difficult linguistic and temporal constraints to which they are accustomed as these

networks expand. More critically, if participation is to be democratic, then NGOs must participate in the development of these networks and be a part of their management. Thus, collaborations are developed for processes internal to a network and for compatibility with processes external to it that are defined by a given summit. It is this particular context that concerns us here. ICTs have only begun to emerge since the middle 1990s that can leverage existing ICTs, such as e-mail and the Web, to help address this complexity.

As an example, Civil Society's drafting processes for the World Summit on the Information Society (WSIS) have been very complicated, tedious, and prone to error. Inputs are received from many NGOs or caucuses composed of NGOs. These inputs may consist of commentary or specific recommendations for language in the consensus document. All of these inputs must be reconciled for inconsistencies, debated, and placed into a structure for the overall document. A tool to assist in this process, making it transparent, democratic, and traceable would make a monumental contribution to the work of NGOs that participate in a transnational context. Complicating the process further is the fact that the NGO community is distributed across multiple languages. There are six official languages in WSIS. Truly democratic debate over document revisions is severely hampered until translations of a draft have been produced.

Legislative Drafting

Legislation in its broadest definition is one of the key products of collaborative agreement processes in transnational civil society work. Legislation here refers broadly to some formal

process of group deliberation that yields an output in the form of a written text. Such texts are usually structured and can take various forms, including bills, resolutions or declarations.

This paper presents a preliminary analysis of requirements for improving transnational, multi-lingual, collaborative work among non-governmental organizations (NGOs) and a review of applicable technologies. The focus of this research is document-based negotiation processes. This is in contrast with both the free form or conferencing communication modalities that are in increasing use in on-line collaboration.

Issues in Multilingual Legislative Drafting

A modest body of literature exists in the area of legislative drafting in multilingual contexts. This derives from processes that have evolved in countries that conduct legislative processes in more than one official language, such as Canada; the Hong Kong Special Administrative Region; and Belgium; as well as International Organizations like the United Nations and the International Labor Organization.

Principles of multilingual legislative drafting

We use the term *legislative system* in a generic sense to refer to a collaborative or deliberative process among multiple entities that has as its goal the production of agreements or statements that are produced in the form of a *text*. This definition is meant to cover the spectrum of legislative processes from parliamentary systems to non-governmental organizations that have agreed to use some type of *rules of procedure* to produce documents. Texts under this definition of "legislative system" might then include, but not be limited to, bills, statutes, declarations or

resolutions, or regulations. Such documents, including drafts versions, are said to be *authentic texts* if they are produced according to agreed-upon rules of procedure.

In *multilingual legislative systems* there is then the designation within the given rules of procedure -- by law in the case of a country or region -- of the *official languages* in which the legislative system may produce authentic texts. The term *language version* will be used here to refer to the instance of an authentic text in a particular language (e.g. Spanish language version).

Two central issues can be identified among societies and organizations referenced in this paper in their implementation of multilingual legislative systems. Decisions about these two issues constrain technological solutions for supporting a legislative system in significant ways. The first and most critical issue is that of defining the status of all of the official languages in the legislative system relative to one another. The second issue is that of defining the principles and processes for interpreting legislation where any authentic text may be represented by more than one language version. These two issues yield a number of corollary issues. All of these are discussed below.

Principle of equality

All official languages must have equal status if participants and "citizens" within the legislative process are to enjoy equal status. That is, one's officially recognized language must not be viewed as subordinate to any other official language within the process. Transitively, one language version of an authentic text must not be viewed as a translation of another language version. All language versions of an authentic text are thus said to be *equally authentic* under such a system.

The *Vienna Convention on the Law of Treaties* (United Nations, 1969) codified these concepts at an international level. Information about multilingual drafting in this paper has been gathered from legislative drafting guides that have been published by the Hong Kong Special

Administrative Region (2001) and Canada (1995); as well as the International Labour Organisation (2001).

Natural language translation issues

The general characteristics of natural languages and the unique characteristics of each language often raise problems in multilingual legislative drafting given the fundamental requirement that equivalence must be obtained between language versions. A complete taxonomy of problems involved in natural language translation is beyond the scope of this paper. Miller (1995) provides an overview. We discuss instead key issues present in the literature on the specific domain of legislative drafting: whether an expression has a direct translation into another language, problems posed by polysemy, and the use of domain-specific terminology.

Any official language likely contains words that have no direct translation into another official language. It may also be the case that a domain-specific term in one language -- such as those found in legal, technical, or scientific literature -- has no direct translation in one of the official languages. These situations have been addressed in some legislative systems through the implementation of multilingual glossaries that attempt to harmonize translations between terms that commonly arise. This has the benefit of not only resolving translation problems, but also of making the overall process more efficient.

It is not uncommon in multilingual drafting to encounter a word in one official language that is polysemous or to find that one official language does not offer an appropriate word that has the same scope of meanings in another official language (Canada, 1995; the Hong Kong Special Administrative Region, 1998). In these cases, one language may require the use of several words to capture the breadth of meaning of its equivalent text in another language. Conversely, what may be required is the choice of a less polysemous word in the former language to constrain the possible meanings in the second.

Sequencing in the drafting process

Because all official languages are to be equivalent, the process of drafting legislation must not privilege one language version over the other. That is, drafting may start in any one of the official languages with translations into equivalent language versions made later. Another possibility that occurs in some legislative drafting systems is that language versions of a text are drafted in parallel in order that drafters can collaborate on finding equivalent meanings. This is known as *co-drafting* (ILO, 2001). The latter approach has the advantage that possible problems in differences across language versions can be dealt with immediately.

In Canada and Belgium both serial drafting and co-drafting are legal options, but it has become customary to use co-drafting. The International Labour Organization's legislation guidelines note that in Belgium drafting often commences in the "mother tongue" of the subject matter expert for the legislation in cooperation with a lawyer producing an equivalent version in the other language (ILO, 2001).

The scope of an authentic text

In multilingual legislative systems, a text is comprised of all of its language versions together. In countries such as Canada or the Hong Kong Special Administrative Region it may not only be the case that language versions are taken to be equivalent. There may be further requirements in legal processes that a text must be considered to include all of its language versions together (Hong Kong, 1998).

Document structure and drafting conventions

Many legislative systems require documents to be produced according to established templates that dictate the types of sections that are allowed in a text, such as chapters, parts, sections, or subsections; how such section types are to be used; and constraints on the way language itself is used. The United States House of Representatives (2004), for example, has established several *styles* by which legislation may be structured. These have subsequently been articulated in the form of XML DTDs or schemas.

Multilingual drafting heuristics and the needs of legislative drafting call for each language version to have the same structure. Structural equivalence refers only to the section level organization of the legislation. It is recognized that sentence level structure between language versions must be allowed to differ due to the diversity of characteristics of each official language. Canadian guidelines, for example, recognize that approaches to reducing ambiguity in legislation differs significantly between French and English (2004, section 2.7).

Glossaries

Glossaries have important functions in legislative drafting in general and in multilingual drafting. Drafters in any language may require access to definitions and usage information about domain-specific terminology. Terminology in one language may not be amenable to direct translation into another language (see Canada, 1995 for examples). For this reason, multilingual glossaries have been proposed and implemented. The Hong Kong Special Administrative Region, for example, has implemented the Bilingual Laws System (BLIS) which has a glossary of legal terms in Chinese and English in addition to full texts of legislation (2001, Annex VII).

Drafting rules or heuristics

A system of drafting rules that specifically addresses multilingualism may be codified within a legislative system. This is the case for the ILO (2001); the Department of Justice in Canada (1995); and the Hong Kong Special Administrative Region (1998). A legislative drafting tool set

should support adherence to such rules. A set of rules has been collected from these sources in table 1.

Table 1. Rules and heuristics for multilingual drafting

Category	Rule/Heuristic	Sources
Principle of Equivalence	• The overriding rule is only to make the substance of all of the language versions equivalent.	ILO
	• Each draft of each language version must be grammatically correct.	ILO
	• Do not use "phrases or expressions that suggest" one language version has a higher status than another.	Hong Kong
	• One language version must not be "forcibly" changed "to fit the particularities of another language."	ILO
Document Structure	• Make the structure of a text identical across each language version with respect to enumerated organizational units.	ILO
	• Allow the structure of lower level, non-enumerated units of text such as paragraphs to diverge across each language version if the characteristics of the natural language require it.	ILO, Canada
	• Allow the syntax to diverge if necessary.	ILO
	• Allow the number of sentences to diverge if necessary.	ILO
Drafting Process	• Draft in the singular in languages where multiple modifiers can lead to ambiguity.	Canada
	• Use paragraphing in languages where it can help to disambiguate the syntax of a sentence.	Canada
	• Structure texts using a hierarchy of enumerated clause types to facilitate parliamentary debate on the text on a part-by-part basis. (This is the normal approach and is not strictly a multilingualism issue.)	Canada
	• Use co-drafting when possible, where persons fluent in each official language collaborate closely to produce language versions in parallel.	ILO, Canada

Background

The period from 1965 to the early 1970s saw the development of a number of technologies that would help lead to the development of the technologies used now by NGOs for transnational collaboration. Packet switching was crucial in the development of data networks, leading to the inception of the ARPANET in the U.S. in 1969 (Press, 1997). Packet switching had the broader impact of enabling broadband communications and the convergence of many media of communication, including data networks, broadcast radio and television, satellite, and cable television. While the ARPANET was not developed for use by average citizens, it was crucial to the evolution of modern mechanisms now used by NGOs for collaboration. It along with the present day Internet fostered the development of a number of technologies that have subsequently become the basis for much of a global information society, including SMTP (i.e. for e-mail), FTP, TCP/IP, and HTTP. Between 1969 and 1990, use of the ARPANET and its technological and organizational offshoots such as CSNET and NSFNET remained largely outside of the realm of average citizens. Mostly academic, government, and industrial scientists had access to these networks (Zakon, 2001).

During the 1980s, the emergence of the personal computer and the establishment of community networks and proprietary dial-up information services allowed the broader public to become familiar with technologies such as e-mail, electronic bulletin boards, chat rooms, and other on-line services. In 1986, the first FREENET was started in Cleveland, Ohio in the U.S. to provide

free access to a virtual community in which people could share information and discuss issues. The Cleveland FREENET was soon followed by the creation of many community networks in North America and around the world. The private sector also began to contribute in a more direct way to the public's increasing exposure to notions of information society during the 1980s. New and existing on-line information services such as CompuServe and AOL began offering dial-up services geared toward personal computer users.[3] These would also soon expand to have global reach.

During this period, NGOs and other entities in the broader civil society became more involved in moves toward an information society. From 1985 to 1986, PeaceNet, GreenNet, and EcoNet were formed to provide networking for peace and environmental activists around the world. In 1987, the Institute for Global Communications (IGC) was formed. It provided services to an increasing number of organizations and other networks over the next decade (Institute for Global Communications, 2002). In 1990, Association for Progressive Communications (APC) was founded by a coalition of civil society-oriented activist networks around the world (Association for Progressive Communications, 2000, pp. 11-15).

Changes in U.S. government policy allowed commercial dial-up access to the Internet starting in 1990 (Zakon, 2001). In 1992, the World Wide Web was released for public use by CERN (European Organization for Nuclear Research) and by 1993 the first general purpose Web browser – Mosaic – was made available to the public (World Wide Web Consortium, 2000). The World Wide Web initiated a rapid increase in Internet usage by government, business, and civil society (Mueller, 1999, pp. 5). The technical foundation provided by the World Wide Web has

subsequently been the basis for an increasing array of tools for communication, knowledge management, and collaboration. Some of these are described in the next section.

Related Work in Collaborative Technologies

The development of tools to support multilingual legislative drafting must draw on the broad technical areas of markup languages and hypertext, computer supported collaborative work (CSCW), and versioning; as well as the more narrowly focused area of research and development around legislative drafting systems. This research also draws from policies and guidelines that have been established by parliamentary and inter-governmental organizations to manage multilingual drafting processes.

Markup languages and hypertext

Markup languages are now associated almost entirely with the World Wide Web; however, markup languages were originally concerned with typesetting. These languages were used first as written annotations in documents to give layout specifications to human typesetters. Eventually these languages evolved into electronic analogs that could be embedded into electronic documents to direct typesetting machines.

It has been noted that legislative drafting methodologies have been influenced in a significant way by the evolution of markup technologies (United States House of Representatives, 2004). The particular typesetting machines that were chosen dictated the languages that were used. In turn, the forms that these languages took guided the development of text editing software to be used by drafters to produce texts with embedded markup. Legislative drafters then evolved organizational conventions and procedures that were, in part, adaptations to the capabilities and limitations of these text editing systems made in attempts to achieve greater efficiencies and ease

of use. For example, numerous keyboard shortcuts have been introduced into drafting systems used in the U.S. House of Representatives.

Markup languages are used today not only for presentation, which encompasses typesetting, but also to provide structural and semantic information about documents. Correspondingly, the styles used to guide the production of legislative documents are often highly structured and the management and use of legislative documents benefits greatly from having access to meta-data -- data about the content data in the document -- that semantic markup makes possible. The style now commonly used in the U.S. Congress to produce bills provides for a hierarchy of as many as eight types of sections, the basic unit; as well as seven higher level types of groupings (United States House of Representatives, 2004). The formats used by many other countries as well as international organizations, such as the United Nations and the International Labour Organization, can be observed to be highly structured as well.

The current generation of markup technologies for these purposes revolves around the XML standard (W3C 2004). XML (for Extensible Markup Language) is designed to enable structural and semantic markup. It can in conjunction with XSLT, an XML-based technology, be used to automatically produce corresponding presentational markup.

Traditional legislative drafting processes make heavy use of cross-references. These take the form of in-text citations to other documents or to specific sections within the same or other documents (e.g. an act being superceded). An example is the following, which is an act of the U.S. Congress mandating a change to an existing act:

> (a) In <<NOTE: 26 USC 3304 note.>> General.--Section 208 of the Temporary Extended Unemployment Compensation Act of 2002 (Public Law 107-147; 116 Stat. 30) is amended to read as follows:
>
> (U.S.C. Pub. L. 108-1)

Thus, a natural part of the application of current markup language technologies to legislative drafting processes has been to make use of hyperlink capabilities to model the various types of cross references in use. For example, applications used for drafting in the U.S. House of

Representatives offer functions to automate the use of internal cross-references within a document (e.g. to refer to section numbers) and for retrieving information from other documents (e.g. other legislation) (United States House of Representatives, 2004). Relevant XML-based recommendations and standards here include XML Pointer, XML Base and XML Linking (see W3C 2003).

Groupware

A tool set to support legislative drafting would fall into the broad category of groupware systems, which Ellis, Gibbs, and Rein (1991, p. 40) defined as:

> computer-based systems that support groups of people engaged in a common task (or goal) and that provide an interface to a shared environment.

In our problem domain, the common tasks include the drafting and deliberations over drafts of a legislative text. The legislative text itself and the computing environment in which it is created and managed constitute the shared environment.

Under the definition given above, a groupware system is typically characterized by its characteristics along three dimensions (Ellis et al. 1991, p. 41):

1. whether or not it allows multiple users to engage in common tasks or to work in the shared environment in real-time;
2. the degree to which it allows or requires users to share common tasks; and
3. the degree to which it allows or requires users to share a common environment.

The combinatorial possibilities, thus, yield the classic class space and time matrix to characterize groupware systems. Examples of how this might apply to legislative drafting systems are given in table 2.

Table 2. Space-time characteristics of possible legislative drafting systems

	Working on common	**Working on common**

	tasks at the same time	tasks at different times
Sharing of space	• Collaborative editing of a shared document in real-time.	• Collaborative editing of a shared document at different times.
No sharing of space	• Collaborative editing of different copies of a document in real-time. • Real-time display of editing process for observers.	• Routing of a document to different users for editing.

Conceptually, a legislative drafting system might need to exhibit characteristics of all four quadrants in table 2. Heuristics for multilingual drafting suggest tools that allow group editing in real-time given that drafters in each official language would ideally work in close collaboration. This type of collaboration might be facilitated by editing of a shared document representing an aggregation of language versions or through no sharing of a common document, but rather communication of results between distributed drafting sites. Legislative drafting may at times lend itself to asynchronous interactions, where editing takes place at different times. One scenario is that of editing one instance of a document in a shared repository, but at different times. The other is that of routing the document to different locations for editing by different users at different times.

Versioning

Vitali (1999) defines versioning as "the management of multiple copies of the same evolving resource, captured at different stages of its evolution" (p. 1). Legislative drafting involves another layer of complexity here in that the "evolving resource" is a legislative text that is typically *consolidated* from sections that were authored independently. From a parliamentary perspective these sections often represent proposed amendments to the text. Multiple copies or

versions of the text then come into existence as each section is added. Further changes to the text may occur in parliamentary processes where text is "marked up" to signify desired edits.

Arnold-Moore (1995) points out that the because existing legislation may be subject to amendment, it may be necessary for proper legal interpretation (e.g. in a court case) to be able to see the state of the text at any *point in time* during its evolution. Furthermore, historical scholars may learn more about the intent of a document's authors and the events surrounding its creation by examining its various versions. A legislative drafting tool set should, therefore, be able to represent the content of a legislative text in such a way that it can be reassembled at any stage in its "evolution."

Several research efforts are relevant here. Versioning methods for hypertext have been studied since at least the 1980s. Vitali (1999) provides a survey of many of these. More recent work has focused on lower level database indexing schemes specifically for XML (Chien, Tsotras, & Zaniolo, 2002). The WebDAV protocol includes mechanisms for supporting versioning in a Web environment. All of these approaches solve essential concurrency control, access, and storage problems for version management at the level of the markup language and, in the case of WebDAV, at the network level. These approaches are, however, domain-independent and do not address the specific needs of consolidation and modeling legislative document structures.

The *EnAct* system (Arnold-Moore, 2002) is designed to address version management specifically for legislative drafting. In particular, it supports the consolidation process and allows legislation to be viewed at different points in time. It makes use of an XML-based change description language that is used to articulate proposed changes to a text, including elimination and substitution of text.

Content Management Systems (CMS) Software

Content management systems has evolved as a distinct class of Web-based software. They have eliminated the need for writing custom backend systems for most tasks involved in managing

content published on the Web. They help to aggregate content into a consistent framework for presentation, as opposed to the tedious task of file-based management. The basic areas of functionality provided by CMS include:

- content aggregation
- metadata management
- authentication
- tracking/monitoring of viewers of the content

Typically content is input into a CMS using a Web client (e.g. through pointing and clicking). However, different modalities of input are provided by CMSs. Not all provide the same modalities. Content might also be input into a given system through:

- e-mail
- Internet chat or Instant Messaging
- news protocols such as NNTP

These systems can be configured to assist in posting of content, taking feedback on content, managing calendars and scheduling, and annotation and limits forms of voting. Some proprietary systems support workflow management, collaboration, and versioning.

These systems usually have modular architectures, which allows new functionality to be added (e.g. calendar, voting, etc).

Popular free CMSs include:

1 PHPNuke

2 PostNuke

3 Zope

Such systems are already used by many NGOs, but they might provide the basis for a more advanced integrated tool set that supports collaborative processes between NGOs.

A wide array of proprietary CMSs are available. See http://www.contentmanager.eu.com.

General Collaboration Technologies

A number of systems have been developed that support various aspects of versioning and collaborative work described above, but which are not specifically designed to support legislative drafting. Notable examples include BCSW (Appelt, 1999) and Hyperwave (formerly Hyper-G) (Andrews, Kappe, & Maurer, 1995). Both systems are not free and open source and, thus, do not meet the criteria that we argue are necessary for supporting civil society. Nonetheless, they provide important models for Web-based content management and collaboration that can be adapted to this domain.

Hyperwave's strengths are in the areas of content and link management. Any applicability to collaboration is, thus, a side effect of strengths in these areas. Content and links are stored separately in a database in this system, which allows greater flexibility in the types of objects that can be assigned links, as well as allowing dangling references to be dealt with more effectively. Hyperwave also provides mechanisms for aggregating content, such as multiple language versions, of a single "document" into one object.

BSCW (Basic Support for Collaborative Work) is a system that has explored the use of HTTP servers and HTML to provide an environment for collaboration. BSCW is based on the concept of workspaces, which are locations where multiple users can collect documents and other types

of objects for joint tasks. Another major feature the system provides is monitoring and notification of events within workspaces, such as document changes. BSCW implements concurrency control and versioning, which are critical for the collaborative use of shared resources.

An alternate approach that is not dependent on a specific set of systems is a protocol-based focus on collaboration. The key example here is WebDAV (Whitehead & Goland, 1999), an open standard for collaboration on the Web. WebDAV articulates an HTTP-based protocol for supporting distributed authoring and versioning. Its goal is to achieve network-based interoperability of tools. This approach opens the possibility for existing applications, such as word processors, to be used for collaborative authoring through minimal extensions necessary to satisfy the protocol. WebDAV provides for concurrency control; the management of "namespaces," which can represent aggregations such as in BSCW; and an approach to extensibility that does not "break" existing applications.

Technologies for Legislative Drafting

Software systems have been used to support legislative drafting since at least the 1960s through the adaptation electronic typesetting software like the Master Typography Program (MTP) used by the U.S. Congress (n.d.). The lineage discussed here is distinct, though not unrelated to, legal retrieval systems (see Arnold-Moore, 1998). Later with the availability of WYSIWYG authoring tools, applications such as Xywrite and Xmetal were integrated into the drafting process of the U.S. Congress through user-level customizations, such as keyboard shortcut definitions, to support the specific needs of legislative drafting.

Since at least the 1990s, efforts began to develop systems that support both the authoring and version management aspects of legislative drafting. Version management in these systems seem to have focused less on the concurrency control features required facilitate collaborative drafting with emphasis being placed on automating the consolidation of legislation from amendments and facilitating point in time retrievals, as described above. A key example is the EnAct system

(Arnold-Moore, 2002) that was developed for the provincial government of Tasmania. In addition, some efforts have been made at addressing multilingualism in the context of legislative drafting. One example is the Bilingual Laws Information System (BLIS) designed for the Hong Kong Special Administrative Region (2001). These systems appear to be closed and proprietary and, thus, not useful for supporting civil society organizations.

For reasons discussed above, legislative drafting technologies have followed the move to XML-based data management. Thus, in addition to XML-based software architectures, such as EnAct, open XML Schemas (and DTDs) have also been developed specifically for legislation. These include open XML-based schemas and style sheets developed for the U.S. congress (n.d.) and the MetaLEX open standard for legal documents developed in the Netherlands (Winkels, Boers & Hoekstra, 2003).

BROAD SYSTEM REQUIREMENTS

The goal of strengthening transnational networks within civil society must include the development of ICT tool sets that solve both the technical and social problems involved in managing formal or semi-formal democratic processes. A number of CMS now exist that might be extended and adapted for this purpose, but no fully functional system as such exists. A critical factor in providing this type of tool is the use of a free software model. Many systems that deal with collaboration and multi-lingual drafting are proprietary.

The broad requirements for a tool set are the following:

- **Management of distributed, democratic processes**: The system should allow collaborating NGOs to easily codify agreed-upon processes and work flows within the network of

participating organizations. The system would then help to manage and monitor these processes. Such processes might include nominating and voting processes (e.g. for people, committees, documents, and referenda); scheduling support for information dissemination by the network; and the editorial and political processes involved in crafting consensus documents.

- **Management of document versioning**: A well-known problem in distributed collaboration is version control management (Benatallah et al., 2003). This is the process of preventing and resolving conflicts in the editing of common documents by multiple actors. Such a system should provide the ability to "lock" documents or sections therein while a certain person or group is working on it. A versioning system would also allow people to visit or "track" older versions and to resolve multiple distributed edits into a common document.
- **Management of linguistic diversity**: Such a tool set must support techniques and processes necessary to produce collaborative documents in different languages. Most often there are multiple official languages in these settings, which frames the minimal expectations for document translations. As a matter of principle, however, a truly democratic transnational network must provide multi-lingual support to any linguistic group, since NGOs must provide interfaces to civil society outside of summit processes.

Translation problems exist on multiple levels. It is obvious that truly democratic debate over document revisions is severely hampered until translations of a draft have been produced such that a whole network of NGOs can participate. Other more subtle problems exist in achieving this, however. Often, translations -- particularly in technical fields -- cannot take place in a linear fashion, from one language to another. Sometimes terms or expressions used

in the drafting language do not have adequate translations in other languages. Thus, a tool set must coordinate simultaneous drafting across languages.

Translation also presents particular requirements with respect to versioning. A great deal of effort is expended in producing any one translation. If further editing takes place in one language, it can be very difficult to identify the parts in the translations that need to be modified. Thus, a tool set must perform cross-language versioning.

- **Development and management of collective knowledge**: In the process of collaboration, such at that in WSIS, a great deal of knowledge is assembled, developed anew, and modified. This happens when individual NGOs or groups of NGOs representing issue areas contribute information or perspectives on a domain being considered by the whole network. A major amount of energy is expended in trying to collect and manage this knowledge in support of the overall process. Currently this takes the form of Web sites that assemble documents and commentaries, archives of e-mail discussions, or domain and language-specific glossaries. Such collections are what Pierre Levy (1997) calls *cosmopedia*: intentional knowledge spaces that are "deterritorialized," yet can be easily located and used for developing a "collective intelligence." The difficulty is that without additional software support, such collections are not easily built, linked and searched without special tools.

A tool set must support knowledge management in a way that is contributory, multi-dimensional, analyzable, and extensible. The design must be geared toward use by average ICT users. Such a system is referred to generically as a *contributory encyclopedia*. The data

model would support the representation of composite documents and would be amenable to both rich forms of hyperlinking and database indexing for efficient searches.

At a high level, a tool set that satisfies all of these requirements would be a modular database-centric document management system, where modules would provide specific functionality outlined above. It would also be "extensible" in that it will be possible for any developer within the community to conveniently add features to its technical design.

Such a system would also ideally be developed under the GNU "copyleft" regime so as to minimize barriers to use by NGOs.

A Preliminary System Design

A tool set would be a collection of client/server applications, operating at the most basic level across the hypertext transfer protocol (HTTP). The clients are usually Web browsers in this case. The planned implementation might consist of Java Servlets operating on the server side. They would store, retrieve, and manipulate content stored in a database.

An alternate implementation scenario that should be explored is to adapt or write a special module for an existing free content management system (CMS) such as PostNuke, PHPNuke, or Zope. This would have the advantages of saving time, allowing the use of other existing CMS modules (e.g. calendar and news), and giving users of the system a much wider support base. These systems all make use of an underlying database management system such as MySQL.

Data Model

The basic type of object managed by the server is a *document*. All documents exist as first class objects in the database server and have unique identity. Documents may be one of two semantically related types: *article, annotation,* or a *ballot.*

Articles: A document of type article corresponds to what would typically be a joint declaration prepared by multiple NGOs or representatives thereof. A document might also include multimedia assets, when deemed necessary, such as images, audio or video. Each article, unlike traditional documents, may be composite in the object-oriented sense, containing other articles. This composite model allows consensus documents -- which are typically structured into number paragraphs -- to be represented in a way where its sections can be managed and updated independently.

Annotations: A document of type annotation is associated with a document of article type. An annotation would be created by participants to provide natural language responses to individual sections of a document. Responses might be commentaries or they may proposals for substitute language.

Ballot: A ballot is a special type of document that allows the recording of evaluative responses to individual sections of a document. The most basic type of response would be a quantitative binary vote -- for or against. The idea, however, is to allow the modeling of different types of evaluations. For example, participants may decide that instead of using ordinal measures, it is preferable to evaluate each section using surveys that allow for

response values that are based on scales (e.g. "strongly approve", "approve", "neutral", "disapprove", "strongly disapprove").

All documents and their constituent objects are semi-structured data objects modeled in conformant XML (Abiteboul, 1997). The semantic relationships between these document types are depicted in *figure 1* below.

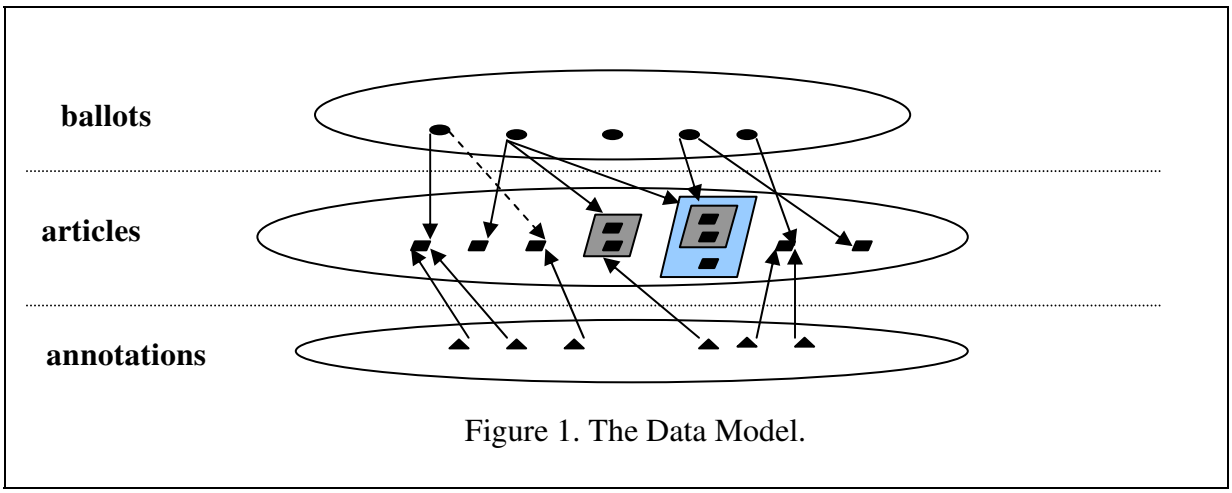

Figure 1. The Data Model.

On an editorial level, the tool set must also support optional editorial layers of articles. Using this feature, a joint NGO declaration could be organized, for example, into one or more collections of articles each developed by a caucus specializing in a particular area. This would entail a configurable workflow management layer that allows each component of such a document to be staged, approved, and collected at different points in the drafting process. A depiction of an example workflow is given in *figure 2*.

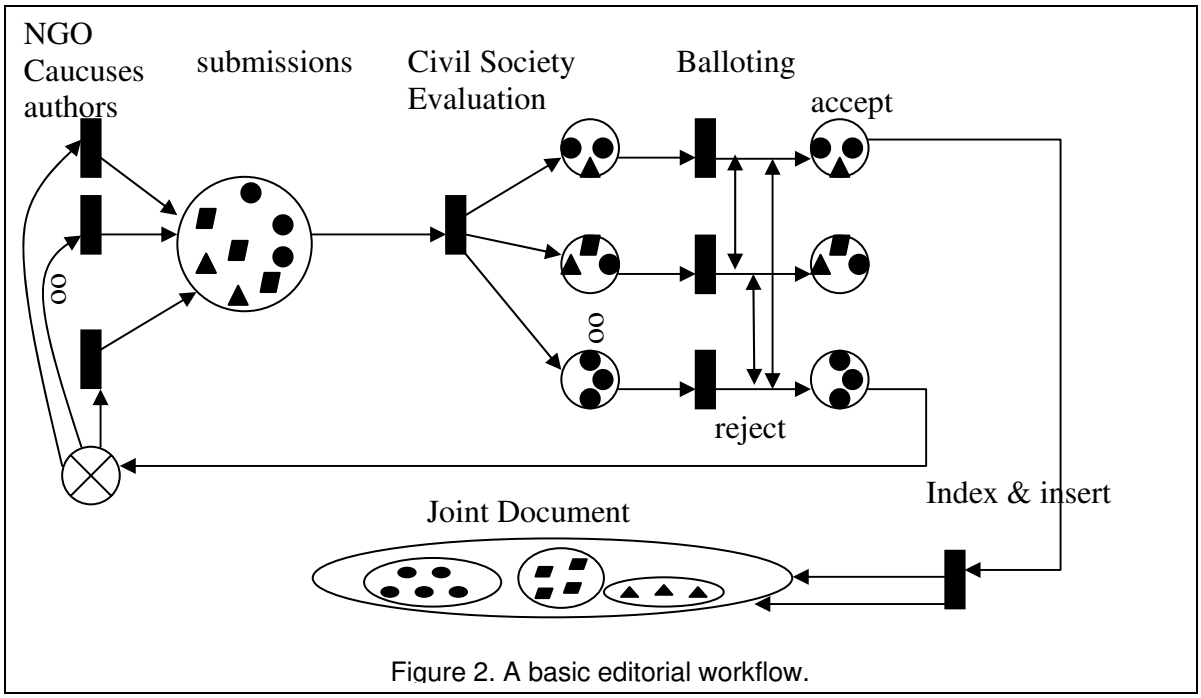

Figure 2. A basic editorial workflow.

Supporting Tools

The system will consist of a set of extensible tools, including user interfaces for authors, editors and browsers, an XML database server, an XML-based workflow management subsystem. The basic configuration is depicted in *figure 3* below. The architecture will be developed around an XML-based database server having transaction management capabilities. A prime candidate for this is the *MySQL* database management system (http://www.mysql.com). The user interfaces will be Web browser-based. The Browser user interface will provide a range of features for locating and retrieving articles in a document. The Management user interface will provide features for performing administrative tasks over a collection of articles. The Authoring user interface will provide a high-level, graphical interface for composing documents. The Document

Manager will handle the assembly of complex documents, document updates, and enforcement of access permissions (Arnold-Moore et al, 2000). The Workflow Manager will automate the routing of documents as part of the editorial process. Workflows will be configurable using specifications written in a language such as the WfMC extension to XML.

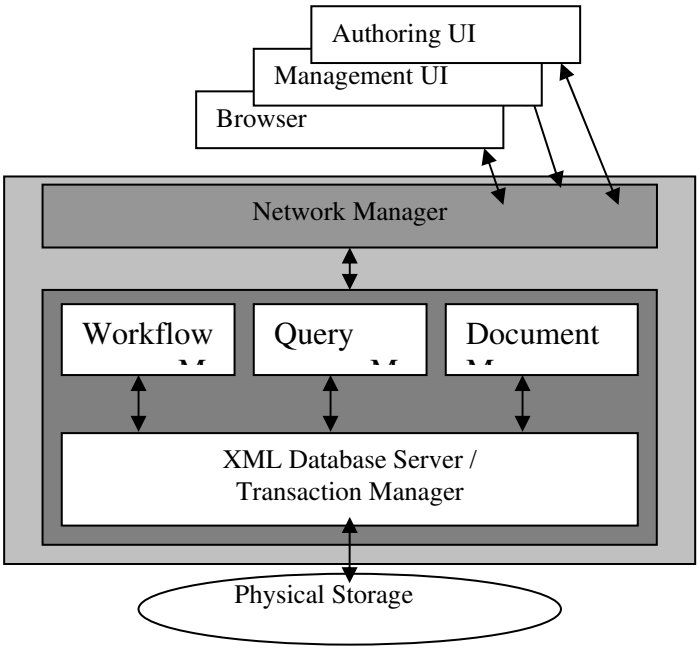

Figure 3. The system architecture

Analytic Tools

Tools would be needed to analyze workflows and the state of processes. They might answer questions about the existence of bottlenecks or how work is being shared across a network of NGOs. In addition, tools that can analyze and compare documents and collections of knowledge

produced by a network would be highly useful.[2] A wide array of statistical methods has been developed for performing quantitative historiography. Such tools might scan a collection of articles and derive totals, averages or trends within them. The encoding of content using structured markup languages such as XML allows the representation of semantic information makes documents amenable to quantitative methods. Several approaches exist for implementing such analytic tools. Such a tool might be implemented as a browser-side plug-in. Another approach would be to implement a stored procedure facility for the database server, wherein analytic functions can be defined and used in the context of query expressions from the Browser/Query user interface.

Conclusions

This paper examined a very specific and crucial computing domain for improving collaboration within civil society networks: technological support for collaborative, multilingual, legislative drafting. The development of legislation, as broadly defined in the first section, is only one mode of production within transnational civil society networking, but one that appears to be gaining in importance as non-governmental organizations continue to participate more fully in international conferences and summits.

The dilemma for civil society organizations is that while a limited number of technologies have been developed to support legislative processes, including drafting, they are proprietary and it appears from our survey that major work remains in addressing the multilingual dimension at a technological level. Free and open source technologies do exist that support various forms of on-line collaboration, but none appear suitable or tailored for supporting the specific needs of legislative drafting, much less the needs of drafting in multiple languages.

[2] In fact, a number of researchers involved in WSIS are performing content analyzes of documents produced in the preparatory process.

Fortunately, the bodies of research in the constituent areas relevant to multilingual legislative drafting and collaborative computing are sufficiently mature to inform the development of free and open tool sets. Well-defined heuristics for multilingual legislative drafting have been derived from experiences in intergovernmental organizations, countries and regions that have multilingual legislative processes. These heuristics can inform the development of preliminary requirements for a tool set to support this form of collaboration. A number of areas within computer science, including CSCW and database systems, have produced well-understood techniques for supporting document versioning, point in time retrieval, and concurrency control for the safe manipulation of a shared document by several authors simultaneously.

The most promising approach to the development of a free and open tool set for collaborative, multilingual drafting will combine the open WebDAV protocol with use of XML-based document technologies. In the aggregate, these technologies will allow a wide array of existing software applications to be leveraged to produce a solution. They are open and flexible enough to create a solution that can evolve. Solutions based on these technologies can and have been ported to many different types of computing platforms.

REFERENCES

Abiteboul, S., D., Quass, J. McHugh, J. Widom, and J. Wiener. The Lorel Query Language for Semistructured Data. (1997) *International Journal on Digital Libraries*, 1(1):68-88, April.

Andrews, Keith, Kappe, Frank, and Maurer, Hermann. (1995). Serving Information to the Web with Hyper-G. *WWW'95.*

Arnold-Moore, Timothy, Fuller, Michael & Sacks-Davis, Ron. (2000) "System Architectures for Structured Document Data," *Markup Languages: Theory & Practice*, Winter.

Arnold-Moore, Timothy. (2002). "Markup for automating legislative processes." *SHIP project 5th Joint Symposium.*Cyberlaw Study Group and Legal Infomatics Study Group. 27th of April 2002. Meiji University in Tokyo, Japan. Retrieved from the WWW in May 2004. http://ship.mind.meiji.ac.jp/lib/sympo5/english.html

Association for Progressive Communications. (2000). *APC Annual Report 2000*. Retrieved November 2002 from the World Wide Web: http://www.apc.org

Boualem Benatallah, Mehregan Mahdavi, Phuong Nguyen, Lionel Port, Quan Z. Sheng, William J. McIver, Jr. (2003). "Pertinence-Centric Document Version Management Scheme." *The 15th Conference On Advanced Information Systems Engineering (CAiSE'03)*. Klagenfurt/Velden, Austria, 16 - 20 June, 2003.

Chien, S.-Y, Tsotras, V., & Zaniolo, C. (2002). Efficient schemes for managing multiversion XML documents. *The VLDB Journal.* 11:332-353.

Dublin Core. (1997). Dublin Core Metadata Element Set, Version 1.1: Reference Description 1999-07-02. <http://dublincore.org/documents/1999/07/02/dces/>.

Hong Kong Special Administrative Region. Law Drafting Division of the Department of Justice. (1998). *A Paper Discussing Cases Where the Two Language Texts of an Enactment are Alleged to Be Different.* Retrieved from the World Wide Web on 11 May 2004. http://www.justice.gov.hk/eng/inpr.htm

Hong Kong Special Administrative Region. Law Drafting Division of the Department of Justice. (2001). *Legislative Drafting in Hong Kong: Crystallization in Definitive Form (2nd Edition).* Retrieved from the World Wide Web on 11 May 2004. http://www.justice.gov.hk/eng/pdf/ldhkv2.pdf

International Labour Organization (ILO). (10 December 2001). *Labour Legislation Guidelines*. Retrieved via the WWW 7 May 2004. http://www.ilo.org/public/english/dialogue/ifpdial/llg/noframes/ch10.htm

ISO/IEC 13250 (1999) Topic Maps: Information Technology Document Description and Processing Languages, December 3.

Hans Klein. (2003). "The World Summit on the Information Society (WSIS): Distilling Lessons from Prior UN Summits" Paper presented at the Society for the Social Studies of Science (4S) Conference.

Kraemer, Kenneth L. and King, John Leslie. (1988). "Computer-based systems for cooperative work and group decision making". *ACM Computing Surveys* (CSUR). Volume 20 , Issue 2 (June 1988). Pages: 115 - 146.

Levy, Pierre (1997) *Collective Intelligence: Mankind's Emerging World in Cyberspace*, (translated from the French by Robert Bononno), Cambridge, Massachusetts: Helix.

Miller, Clark A.. (2003). "Science and Global Sustainability: Where Do We Go Next?", , U of Wisconsin-Madison." Paper presented at the Society for the Social Studies of Science (4S) Conference.

Miller, G. (1995). WordNet: a lexical database for English. *Communications of the ACM*. Volume 38 , Issue 11 (November 1995). Pages: 39 - 41.

Mueller, M. (1999). ICANN and internet governance sorting through the debris of 'self-regulation'. *Info*. vol.1, no.6, December, 497-520. Retrieved in November 2002 from the World Web: http://www.icannwatch.org

Ó Siochrú, Seán (2003, November) GLOBAL GOVERNANCE OF INFORMATION AND COMMUNICATION TECHNOLOGIES: IMPLICATIONS FOR TRANSNATIONAL CIVIL SOCIETY NETWORKING. Report to the IT Governance and Civil Society Research Network, Information Technology and International Cooperation (ITIC) Program Social Science Research Council. http://www.ssrc.org

Press, L. (1997). Seeding Networks: The Federal Role. January/February 1997. Retrieved in November 2001 from the World Wide Web: http://www.isoc.org

Schreibman, Susan, Amit Kumar and Jarom McDonald. (2003). "The Versioning Machine." *Literary and Linguistic Computing*. Volume 18, Issue 1, April 2003: pp. 101-107.

Sperberg-McQueen, C.M.. and Burnard, L. (eds.) (2002). TEI P4: Guidelines for Electronic Text Encoding and Interchange. < http://www.tei-c.org/P4X/>.

United Nations. International Law Commission. (1969). *Vienna Convention on the Law of Treaties*. Adopted 22 May 1969. Entry into force on 27 January 1980. Retreived from the World Wide Web on 2 June 2004. **http://www.un.org/law/ilc/texts/treaties.htm**

U.S.C. Pub. L. 108-1. (Jan 8, 2003). "An Act: To provide for a 5-month extension of the Temporary Extended Unemployment Compensation Act of 2002 and for a transition period for individuals receiving compensation when the program under such Act ends." 108th Congress. Jan. 8, 2003 - [S. 23]. Retrieved from the WWW on 25 May 2004. http://frwebgate.access.gpo.gov

United States House of Representatives. (2004). Drafting legislation using XML. Retrieved from the WWW on 7 May 2004. http://xml.house.gov

Vitali, Fabio. (1999). "Versioning hypermedia." *ACM Computing Surveys*, Vol. 31, Number 4es, December 1999.

W3C (2000a) *XML Schema Part 0: Primer*, W3C Working Draft, 7 April 2000. http://www.w3.org/XML/Schema

W3C (2000b) *XML Schema Part 1: Structures*, W3C Working Draft, 7 April 2000. http://www.w3.org/XML/Schema

W3C (2000c) *XML Schema Part 2: Data Types*, W3C Working Draft, 7 April 2000. http://www.w3.org/XML/Schema

W3C (2003). W3C XML Pointer, XML Base and XML Linking. http://www.w3.org/XML/Linking

W3C. (2004.). *Extensible Markup Language (XML) 1.0* (Third Edition) W3C Recommendation 4th February 2004, François Yergeau, Tim Bray, Jean Paoli, C. M. Sperberg-McQueen, Eve Maler. Retrieved for the WWW on 25 May 2004 http://www.w3.org/TR/2004/REC-xml-20040204

W3C. (1999). *XSL Transformations (XSLT) Version 1.0.* W3C Recommendation 16 November 1999. Retreived from the WWW on 25 May 2004. http://www.w3.org/TR/xslt

Whitehead, Jr., E. James & Goland, Yaron Y. (1999). WebDAV: A network protocol for remote collaborative authoring on the Web. *European Computer Supported Cooperative Work conference*. Copenhagen, Denmark, 12-16 September 1999.

Winkels, Radboud. Boers, Alexander, and Hoekstra, Rinke. (2003). *XML Europe 2003.* 5-8, May 2003, London, England.

Zakon, R. H. (2001). Hobbes' Internet Timeline v5.4. Retrieved in November 2001 from the World Wide: http://www.zakon.org