

Jon Bing and the History of Computerised Legal Research – Some Missing Links

GRAHAM GREENLEAF*

[A chapter in Olav Torvund and Lee Bygrave (Eds) *Et tilbakeblikk på fremtiden* ("Looking back at the future"), Institutt for rettsinformatikk, Oslo, 2004]

Introduction

The two bookends for this paper are Jon Bing's groundbreaking *Handbook of Legal Information Retrieval*¹ from 1984, and twenty years later his 2003 article "The Policies of Legal Information Services: A Perspective of Three Decades".² Almost exactly half way between these two surveys, the use of the Internet for the dissemination of legal texts "took off".

The vision uniting these two works is that of an integrated national legal information service. It was and is a grand vision. It is one that Jon Bing realised before almost anyone else, has charted its history more than anyone else, and continues to argue for with passion and elegance. Yet it may be a somewhat flawed vision, one that now needs a new element added to it, an element that could not easily have been realised in 1984.

Jon Bing has done more than any other person to chart the international development of legal information retrieval systems over the whole of their history, so it is worthwhile to consider his perspectives when looking forward in 1984 and looking back in 2003.

The view from 1984

In 1977, Bing and Harvold published *Legal Decisions and Information Systems*,³ one of the earliest theoretical analyses of legal information retrieval.⁴ Parts of that book were expanded and updated in the 1984 *Handbook*, but the latter went further by including a survey of all legal information retrieval systems and projects in the world at that time. Part I examined "Legal decisions and communication processes", Part II was an exhaustive examination of text retrieval, surveying all research on that subject to date, and an "[i]nternational survey of legal information retrieval" which reviewed developments in 25 countries, plus a handful of research projects related specifically to law. It was a groundbreaking contribution in 1984, and has not been matched to this day.

It would be very interesting to assess how many of the systems Bing described in 1984 have survived the last twenty years in one form or other. The one Australian system

* Professor of Law, University of New South Wales; Co-Director, Australasian Legal Information Institute (AustLII) and World Legal Information Institute (WorldLII); Director, Baker & McKenzie Cyberspace Law and Policy Centre, University of New South Wales.

¹ North Holland, 1984; now available at <<http://www.lovddata.no/litt/>>; Bing was the principal editor of the *Handbook* but contributions were also made by Tove Fjeldvig, Trygve Harvold and Robert Svoboda.

² In Lee A Bygrave (ed), *Yulex 2003*, Institutt for rettsinformatik / Norwegian Research Centre for Computers and Law, Oslo, 2003, pp 37–55.

³ Jon Bing and Trygve Harvold, *Legal Decisions and Information Systems*, Universitetsforlaget, Oslo, 1977.

⁴ Colin Tapper, *Computers and the Law*, Weidenfeld & Nicolson, London, 1973, was a major survey of text retrieval systems in law published four years earlier.

described, the pioneering Commonwealth Attorney-General's SCALE system, managed to survive as a system of national scope (as SCALEPlus on the Internet) until 2003 when it abandoned all of its databases except Commonwealth legislation. In the United Kingdom, the independent services, Eurolex and Infolex, that were prominent in 1984 have disappeared, Prestel lives but has been acquired by the Thomson group, and the 1984 alliance between Butterworths and Lexis has transformed into their being the same company. The publicly owned systems that dominated the 1984 description of the USA, FLITE and JURIS, are now of no importance, but the only two commercial systems to rate a mention in 1984, Lexis and WestLaw, are now two of the three international legal behemoths, together with Kluwer (which featured in the 1984 chapter on Holland). We need a Jon Bing to analyse the technology and politics of the last twenty years, but there is no doubt that one of the leitmotifs of any such history would be relentless global expansion of the "big three" and the complementary disappearance of independent national commercial legal publishers.

What wasn't obvious in 1984

Reading the *Handbook* with the benefit of hindsight, we can see that there were a handful of revolutionary developments which would completely change the nature of legal information retrieval during the next decade, few of which are hinted at in the *Handbook*. Four years later, I co-authored an *Australasian Computerised Legal Information Handbook*⁵ (with a foreword by Jon Bing), and my co-authors and I did not anticipate any of these revolutionary changes either. The *Handbook* did not attempt to predict the future, and should not be criticised for failing to do so.

Four of these seismic events in the decade 1984-94 were the CD-ROM, hypertext, relevance ranking of search results, and free access to law. I shall examine each of these in turn, and what Jon Bing has to say about them from the perspective of 2003 in his recent article.

The CD-ROM

At the time of publication of the *Handbook* – which coincided almost exactly with Apple's launch of the Macintosh and not long after IBM launched its PC – it would not have been obvious to Bing that the 1980s would become what he now describes as "the decade of distributed computing". By the mid-80s, personal computers were working their way onto lawyer's desktops, and the previous prerogative of librarians was starting to be reconceived as "desktop research". Bing, always pragmatically minded, often quipped that the greatest impediment to legal retrieval systems was the distance between a lawyer's desk and the library, the extreme case of (to use the *Handbook*'s terminology) "availability failure".

As Bing points out, what made the difference for legal information retrieval in the decade 1985-95 was the invention (prototyped in the early 80s) and commercialisation of the CD-ROM as a means of providing a legal information retrieval system to the technologically savvy lawyer's desktop. Perhaps there were not many of these to begin with – but not many were using dial-up services either.

⁵ G Greenleaf, A Mowbray and D Lewis, *Australasian Computerised Legal Information Handbook*, Butterworths, Sydney, 1988.

In Australia, the CD-ROM rapidly replaced dial-up services as the legal information service of choice around 1990, and allowed a number of new entrants into computerised legal publishing, in contrast to a monopoly service for online information. The impact of the CD-ROM on each country's legal information services must have been different, but no-one has written an equivalent of Bing's *Handbook* for the decade 1984-94 to examine this in detail.⁶

Bing examines why the CD-ROM made such an impact: large storage (650 MB) compared with magnetic disks at the time; availability on your own PC along with other office tools; its marketability as a physical object; and its adaptability to boutique legal collections rather than "one size fits all".

A couple of additional factors, which are easy to forget in the age of web interfaces and broadband connections, are that the interfaces of CD-ROM retrieval systems were usually so much more inviting than the line-based online interfaces, and the speed of retrieval was usually blindingly fast in comparison.

In any event, dial-up retrieval systems were walking corpses, bleeding money, in at least some countries (such as Australia) by the early 1990s, unless their CD-ROM divisions kept them alive. As Bing says,

The new paradigm was that of a large number of self-sufficient islands. The idea of the integrated legal information service was perhaps not lost, but certainly downgraded.

Hypertext and decentralisation

Bing explains in 2003 how the Internet and web browsers created a radically different online world from the one he describes in 1984 in the *Handbook*:

... the structure of this communication network was very different from what had been common at the end of the 1970s. At that time, one would draw a centralised network around a main hub, linking all users to this facility; this made communication between users possible, but only through the central hub. The Web and Internet was more of a distributed network, and node could communicate to any other node, and it did not have such an obvious hierarchical structure.

This, he argues, brought forth both possibilities and problems in the development of national legal information services:

One of the fascinations with the new technology was the ease and low cost of establishing home pages, enabling institutions to present themselves and to provide information to the external world. Rather than using the technology to collect information from many sources, presenting them in a coordinated and integrated way, a profusion of initiatives was taken – each court, each agency, each institution presented their own site to the public. Initially, therefore, the tendency to fragment was further emphasised. The user, who wanted to search for national sources, was presented with a rather differentiated, if not confused situation ...

Bing is right to emphasise the radically decentralising nature of the Internet and the web, but it is surprising that hypertext, the key technology involved in these development, only rates a couple of passing mentions in the 2003 review: the comment that "hyperlinking across [CD-ROM] disks was awkward"; and the observation that large numbers of hyperlinks can be generated and managed if legislative and other texts are normalised at the time of their creation.

⁶ This would be a good topic for a PhD. Bing's doctoral thesis was on legal information retrieval.

Hypertext did not feature in the 1984 *Handbook* because its practical implementation is usually only regarded as commencing with Apple's HyperCard in 1987.⁷ The 1984 title was *Handbook of Legal Information Retrieval* and that says a lot in itself. Searching (retrieval), whether over the full (authentic) text or over representations of that text (abstracts, index terms, titles etc), was the only technology that was given any consideration in 1984.

I think that hypertext has been one of the few revolutionary developments in computerised legal research. It is, of course, the essence of the world-wide-web. The ability to follow hypertext links from any document on the web to any other document – or even to a place within a document – was a breathtaking leap from its predecessor, Gopher, which was essentially only a matter of being able to connect to the front page of any Gopher site from any other Gopher site.

Insofar as law is concerned, awareness of the importance of hypertext pre-dated the Internet, and was an important part of the CD-ROM revolution that Bing charts. In Australia, the first significant commercial CD-ROMs for law, published by Computer Law Services in about 1990, contained the Commonwealth legislation. They had a reasonable search engine, but what made them so attractive to use was their excellent hypertext, not just for tables of contents, but also for hypertext links from any defined term in a statute to its definition. They used a US hypertext engine and their CD-ROMS were commercially successful until the web and AustLII meant it was no longer viable to sell cases and legislation with relatively little value-adding.

Our own DataLex Project research had developed in 1990 a hypertext engine⁸ and a demonstration dial-up online system combining hypertext and text retrieval.⁹ From then on, we argued that the densely cross-referenced nature of legal materials made them hypertextual in nature, and that the future of computerised legal research was largely to do with finding intelligent ways to combine the strengths of full text searching with the strengths of hypertext (and inferencing systems, but that is another story).¹⁰ Our DataLex Workstations¹¹ from 1991-95 (a short-lived commercial enterprise) were based largely on hypertext. Other researchers in Canada¹² and elsewhere were developing theories and approaches to hypertext and law.

The earliest large-scale web-based law systems, Cornell's Legal Information Institute and AustLII, were successful largely because their complex use of hypertext allowed users with no expertise in searching to find the documents they wanted and to navigate within and between them in complex ways. AustLII had millions of automatically

⁷ As is well known, most of its principles can be traced back at least to Ted Nelson in the early 1960s, and (in a non-computing context) to Vannevar Bush's article "As We May Think" in the 1940s.

⁸ Called HYPE, written by Andrew Mowbray.

⁹ These were developed for a demonstration at the "Laws of Australia" Conference, Sydney, October 1990.

¹⁰ G Greenleaf, A Mowbray P van Dijk, "Representing and using legal knowledge in integrated decision support systems – DataLex WorkStations", *Artificial Intelligence and Law*, 1995, vol 3, nos 1–2, pp 97–124.

¹¹ G Greenleaf, "The Privacy Workstation", *International Yearbook of Law, Computers and Technology*, 1992, vol 6, pp 177–196; G Greenleaf, A Mowbray and A Tyree, "The DataLex Legal Workstation – Integrating tools for lawyers", *Journal of Law & Information Science*, 1992, vol 3, no 2, pp 219–240.

¹² <<http://www.worldlii.org/catalog/270.html>>.

inserted hypertext links from its first release in 1995,¹³ and Cornell's US Code was always heavily and attractively hyperlinked.

Now, of course, we just take for granted the tricks of browsing complex bodies of legislation, or vast catalogs of general websites like Yahoo!, or law-specific catalogs like the WorldLII Catalog not simply because the conventions of web design are more or less observed, but mainly because it is so inherently easy. It is much easier than the enormous difficulties that users faced when using the search engines described in the 1984 *Handbook*.

Yet even text retrieval was about to get easier. And the fragmentation Bing describes was prevented from destroying the Internet.

Spiders and relevance ranking of search results

I recall the first time Jon Bing visited the University of New South Wales, in 1985, and that he was very keen to obtain from our Library a copy of an article in the *Communications of the ACM* by Blair and Maron.¹⁴ They had concluded that it was difficult to see why anyone would bother to use full text retrieval systems, in an experiment concerning use by lawyers of a litigation support database using Boolean retrieval. To the great surprise of the lawyers concerned, it was found that the value of recall was only 0.2 (20%), whereas precision was 0.75 (75%). They believed that they were retrieving more than 75% of all relevant documents, but in fact were only retrieving 20%. Blair and Maron concluded that the main assumption underlying automatic indexing was incorrect, namely that the likely users of these systems can "foresee the exact words and phrases that will be used in the documents they will find useful, and only in those documents". In other words, they question whether, in practice, likely users can capture concepts in word-occurrence combinations. If not, it is irrelevant whether it is possible to do so in theory. They suggest that if high recall is desired, manually assigned index terms ("intellectual indexing") is necessary. If it demonstrated anything,¹⁵ their experiment demonstrated the "trade-off" between precision and recall when Boolean retrieval systems which only had an "identity function" caused users to "AND" multiple search terms in order to avoid relevant items being swamped by irrelevant ones.

How peculiar this discussion must sound to someone who did not suffer earlier generations of text retrieval systems and who has grown up using Google on a daily basis. We are now routinely astonished by both its demonstrable precision (where the thing you most want is so often at the top of the list of search results) and its apparently high recall (though appearances can be deceptive, which was a good point of Blair and Maron's article). Performance of modern search engines for legal retrieval needs the analysis of a lot more Jon Bings, but there is no doubt that information retrieval has

¹³ For details and illustrations, see G Greenleaf, A Mowbray, G King and P van Dijk, "Public access to law via internet: the Australasian Legal Information Institute", *Journal of Law & Information Science*, 1995, vol 6, no 1, pp 49-69.

¹⁴ D Blair and M E Maron, "An evaluation of retrieval effectiveness for a full-text document retrieval system", *Communications of the ACM*, 1985, vol 28, issue 3, pp 289 *et seq.*

¹⁵ There were many reasons to criticise Blair and Maron's experiment. For a brief discussion, see Greenleaf, Mowbray and Lewis (1988) (relevant extract available at <<http://www2.austlii.edu.au/cal/guides/retrieval/handbook/index-3.html#Heading61>>, and further discussion at <<http://www2.austlii.edu.au/cal/guides/retrieval/retrieval-3.html#Heading29>>).

changed dramatically, and that these days novices will obtain results that ten years ago expert searchers would have found difficult to achieve. Why is this?

Jon Bing explored many possible solutions to retrieval difficulties, and identified one partial solution as the combination of authentic texts and abstracts.¹⁶

The principal solution to the problem identified by Blair and Maron was already found in Bing's 1984 *Handbook*, Chapter 4 (entitled "Anatomy of Text Retrieval") which included a lengthy survey of new research into the "nearness function" and how it presumed criteria that "may be used to rank documents in such a way that probable relevant documents will be ranked high". This was contrasted with the "identity function" of retrieval systems (using the *Handbook's* terminology) based on Boolean logic. Bing concluded:

In the last years, several researchers have been developing methods combining Boolean retrieval with ranking. This will increase precision and make the users more satisfied with the results.

He was right, and our reasonable levels of satisfaction today with search engines is courtesy of one of the most "user friendly" developments in legal information retrieval: relevance ranking of search results, optionally combined with Boolean searching (usually on the "Advanced Search" page).

Effective commercial implementation of relevance ranking post-dated the *Handbook*¹⁷ but pre-dated the widespread use of the Internet, two of the most notable early examples in the dial-up era being the 1993 introduction of Westlaw's WIN (Westlaw Is Natural)¹⁸ and Lexis' Freestyle. The state of research in the pre-web era, a decade after Bing's *Handbook*, was summarised by Turtle (1995),¹⁹ one of the developers of WIN.

Relevance ranking was so important because it provided a partial solution to the perennial problem of Boolean retrieval systems: the trade-off between precision and recall, and possibly even the inverse relationship between the two. Of the many varieties of relevance ranking systems, the earliest used two principal measures of word occurrence significance, inverse document frequency²⁰ and within-document frequency.²¹ The weighting to be given to an occurrence of a term in a particular document could be as

¹⁶ Bing argued that the debate about retrieval over full text versus retrieval over abstracts etc was "a thing of the past", at least insofar as commercial retrieval systems are concerned. The experiments of Tove Fjeldvig concerning a database of decisions of the Norwegian Central Tax Authority found that abstracts alone had the lowest retrieval performance (including both precision and recall), full text the next best, and a combination of both abstracts and full text the highest performance. See J Bing, "The law of the books and the law of the files" (Part 1), *Computers and Law*, 1987, vol 54, pp 31 *et seq.*

¹⁷ Bing notes that Ejan Mackaay proposed such a system in Quebec (DATUM II) but that it was not implemented (Bing 1984, pp 320–321).

¹⁸ Sheilla Desert, "Westlaw's Natural v. Boolean Searching: A Performance Study", *Law Library Journal*, 1993, vol 85, pp 713–742.

¹⁹ Howard Turtle, "Text retrieval in the legal world", *Artificial Intelligence & Law*, 1995, vol 3, nos 1–2, parts 5.5 and 5.8.

²⁰ Inverse document frequency is the number of occurrences of a term in whole database divided by the number of words in the whole database. This means (very roughly) that terms which occur in relatively fewer documents in the whole collection are given greater weight (also called the "discrimination value" of the term).

²¹ Within-document frequency is the number of occurrences of a term in a document divided by the number of words in the whole document. Therefore (again in rough terms) a word that occurs a lot in a short document gets a high score on this criterion.

simple as the product of these measures for that term. The measure of relevance of the whole document could then be as simple as the sum of the weights of each occurrence of each search term in the document.²²

In the web environment relevance ranking was just as useful, though few sites had search engines that could implement it in the early years,²³ and not many held a sufficient amount of data for retrieval performance to be a major problem. The fragmentation of data identified by Bing was the real problem that first had to be solved: how to search across multiple web sites simultaneously?

The development of the web spider, or web robot, enabled the creation of Internet-wide search engines, and first came to large-scale public attention with the launch of Alta Vista in early 1996. Alta Vista also used relevance ranking of its search results, based on standard techniques of inverse document frequency. This was probably the first time most people had any experience of a search engine that used relevance ranking.

Since then there has been an eight-year battle between Internet-wide search engines for market share, with Google the clear winner at this stage. Google's success seems to have been a combination of various factors.²⁴ The power of its web spiders to simply index more of the web than its competitors has been one major factor. Another major factor has been its innovation in the development of relevance ranking of results, its proprietary PageRank algorithm which (among other things) measures the number of links that point to a page as one measure of the page's significance. Google's PageRank is a clear example of not only the significance of relevance ranking in the history of computerised research, but also of the importance of the intelligent integration of hypertext and retrieval techniques: without the billions of hypertext links that Google is able to find on pages across the Internet, its ranking system would not work.

For those trying to build legal information systems on the web, both relevance ranking and web spiders are essential. Very large legal web sites like CanLII, BAILII or AustLII would be far more difficult to use if they could not offer to the general public an "any of these words" search with relevance ranked results, or a Boolean search with relevance ranked results to more experienced users. This is part of the answer to the question Bing poses in 2003: can the same legal information service serve the needs of both professional and lay users? To the extent that it can, much of the thanks must go to the development of relevance ranking.

For legal researchers, Google and its competitors are as much of a boon to legal researchers as anyone else, provided one is trying to find known documents. Yet the fact that legally-related materials probably constitute less than 1% of what can be searched on Google and the like presents problems of precision when one is trying to search for terms that may have many uses in non-legal contexts. For this reason, some legal services like WorldLII²⁵ and Droit Francophone²⁶ employ their own web spiders which they target solely at law sites. WorldLII has also developed a "Law on Google" service which

²² See Turtle, *op cit*, for a detailed analysis of the many versions proposed as at that date.

²³ AustLII, for example, only developed its own relevance ranking version of the SINO search engine in 1997

²⁴ A clean and simple interface is one.

²⁵ WorldLII Catalog and Websearch at <<http://www.worldlii.org/catalog/270.html>>.

²⁶ <<http://portail.droit.francophonie.org/>>.

translates a user's SINO searches into Google's syntax and then adds a list of "law words" to the search to attempt to limit the results to law-related pages.

The web spider has not yet solved the problem of fragmentation that Bing stresses, but better development in future of law-specific web spiders, and more sophisticated ways of extracting only the legal content from services like Google, may reduce it to manageable levels.

Free access to law

In the 1984 *Handbook*, free access to legal information was not even an issue for discussion. Online legal information systems were mainframe-powered beasts that required feeding by governments or commercial users with continuous supplies of cash. The usage costs of Australia's CLIRS system at one point reached A\$720 per hour for some databases. The *Handbook* notes connection fees of DEK 495 per hour (Denmark), 5000 Belgian francs per hour, and so on.

The position did not change much during the short CD-ROM era from the late 80s to mid-90s. A number of factors weighed against CD-ROMs ever becoming a vehicle for free access to law. Relevant factors at that time (now changed completely) were the high cost of CD-burners, and the fact that most users outside professional legal offices or University libraries would not have CD-ROM drives. One of the most important factors was the distribution and updating problem: sending out pieces of plastic to users, and then sending out frequent replacement pieces of plastic, is expensive and not something anyone would want to do other than on a commercial basis. In contrast to the Internet, there is also no community involved: no network through which potential new users can find out what is available, and no network through which they can give the feedback necessary to keep free access facilities responsive to user needs and aware that they had an appreciative audience.

The Internet provided a viable, almost cost-free, mechanism for delivery of legal information (users absorbed most of the telecommunications charges), and a network through which some type of user community could be involved in communication about what was provided. As Bing notes, "[i]t was tempting to set up alternative services competing with the older" commercial services, once the web arrived.

True, but someone had to have the idea that free access to legal information was important enough, and as Jon Bing explains, the first people to articulate this and to do something serious about it were Peter Martin and Tom Bruce when they established the Legal Information Institute at Cornell University, and particularly their early realisation (obtained through feedback on the web) that a very large part of their audience was the general public, who wanted access to legal information. Their ideas and model were the seeds of the free-access-to-law movement. Theirs was an integrated legal information institute ("LII" as the generic became known) since it included legislation, caselaw and some secondary content, but it was only "national" in that it focussed on the US federal jurisdiction, not that it attempted to provide coverage of all essential national legal sources. Full-scale "national" LIIs were the next step in the evolution of the Cornell model, in the form of AustLII, BAILII, CanLII etc.

It is still surprising that the Internet made possible a viable third contender in the development of Bing's "integrated national legal information system", but free access

LIIs are increasingly able to hold their own against both integrated government systems and the comprehensive systems of the larger legal publishers.²⁷

Ten years after the formation of the LII (Cornell), a global coalition is forming to foster and help spread the ideals of free access to law across the world. In October 2002, free access Legal Information Institutes from around the world,²⁸ meeting in Montreal at the 4th *Law via Internet Conference*, made a Declaration²⁹ as a joint statement of their philosophy of access to law. The essential elements of the Declaration can be summarised as:

- A demand that governments do not exercise any monopoly control over “public legal information”, and that they make it available to independent parties who wish to republish it. This is discussed further below.
- Agreement between legal information institutes to cooperate with each other.
- Agreement between legal information institutes to adhere to the principles of the Declaration.

The emphasis is on the support for free access. The Declaration does not require participation in any networks, but does encourage such participation. At present, all of the LIIs collaborate in networks, the largest of which is the World Legal Information Institute (WorldLII),³⁰ which Bing mentions but which has now been joined by Droit Francophone,³¹ a French language portal into the content available within and outside the LIIs.

Once is not enough – Public policies supporting free access

Jon Bing discusses (and praises) the development of Legal Information Institutes, but I do not think the full implications of the free-access-to-law movement have yet been taken fully into account in his approach to the development of national legal information systems.

Before discussing the model for an integrated national legal information system advanced in Bing’s 2003 article, it is worth first exploring what the Declaration on Free Access to Law might entail when it states:

Independent non-profit organisations have the right to publish public legal information and the government bodies that create or control that information should provide access to it so that it can be published.

To amplify that policy, we might ask whether it is sufficient if governments provide free access to law through government-run websites. AustLII has argued since 1995 that this

²⁷ For example, according to the ranking service Hitwise (<<http://www.hitwise.com.au>>) in April 2003, AustLII had almost 30% of all legally-related web traffic in Australia (excluding immigration visa sites), exceeding the total of the top 7 government legal sites combined, and exceeding the total of the top 6 commercial legal publishers combined.

²⁸ At the Montreal and Sydney meetings, the participating LIIs were the LII (Cornell), BAILII, AustLII (and WorldLII), LexUM (CanLII, and Droit Francophone), PacLII, HKLII and SAFLII, plus other organisations from the Caribbean, the Pacific and elsewhere.

²⁹ See <<http://www.worldlii.org/worldlii/declaration/>>.

³⁰ <<http://www.worldlii.org/>>.

³¹ <<http://portail.droit.francophonie.org/>>.

is not enough³² and that official bodies should accept that they have seven obligations in the provision of essential legal information if they are to give optimal support to the rule of law and other values:

- 1 *Provision in a completed form*, including additional information best provided at source, such as the consolidation of legislation, and the addition of catchwords (index terms) or even summaries to cases.
- 2 *Provision in an authoritative form*, such as use of court-designated citations for cases and (eventually) use of digital signatures to authenticate the versions distributed.
- 3 *Provision in the form best facilitating dissemination*, which should always now mean in electronic form, should in most cases be possible by email or more sophisticated forms of data delivery, and should be possible in a form facilitating conversion.
- 4 *Provision on a marginal-cost-recovery basis to anyone*, so that governments do not attempt to profit from the sale of public legal information, thereby creating artificial barriers to access to law.
- 5 *Provision with no re-use restrictions or licence fees*, subject only to such minimal restrictions as are necessary to preserve the integrity of published data³³.
- 6 *Preservation of a copy in the care of the public authority*, so that an archive of the data is preserved to enable greater competition whenever a new entrant wishes to publish the data, a whether or not the public authority publishes the data itself.
- 7 *Non-discriminatory recognition of citations*, so that Court-designated citations are not removed from “reported” cases, ending the privileged status of citations of “official” reports.³⁴

A corollary of these propositions is that it is not sufficient for official bodies only to publish essential legal information for free access on their own web sites. Provision to other publishers (including both LIIs and commercial publishers) is also necessary for sound public policy, and is more important than official self-publication. Such dissemination is necessary to ensure that free-access is not second-rate access.

The spirit of these AustLII public policy principles is now reflected in the Declaration on Free Access to Law, but a more detailed elaboration of what “full and free access” implies is also worth undertaking if the full implications of the free-access-to-law movement are to be understood. These wide-ranging propositions cannot be explored fully in this paper, but this summary is sufficient for the purposes of our further discussion here.

³² See G Greenleaf, A Mowbray, G King and P van Dijk, “Public access to law via internet: The Australasian Legal Information Institute”, *op cit*; G Greenleaf, A Mowbray and G King, “New directions in law via the internet – The AustLII Papers” *Journal of Information, Law and Technology* (JILT), 1997, issue 2, at <http://elj.warwick.ac.uk/jilt/issue/1997_2>; and G Greenleaf, “Free the Law: How the Australasian Legal Information Institute (AustLII) Achieved the Free Availability of Legal Information on the Internet”, *Journal of Information, Law and Technology* (JILT), 2000, issue 1, at <<http://www.law.warwick.ac.uk/jilt/00-1/transcript.html>>.

³³ Poulin takes a similar position: “In order to be truly open, publication must not be restricted by any conditions as to subsequent use or republication of the documents, except for conditions needed to ensure accuracy”. See Daniel Poulin, “Open Access to Law in Developing Countries”, 5th Law via Internet Conference, AustLII, Sydney, November 2003 (forthcoming in *Journal of Information, Law and Technology* (JILT)).

³⁴ This seventh element was not included in our original list in 1995, but the importance of citations has subsequently become more obvious.

Jon Bing's model for national legal information systems

Bing describes the objective of his 2003 paper as

to indicate a strategy for the development of an integrated, national legal information service which does not presume a certain technological infrastructure, and which may grow and adapt to the changing social and technological context.

It is particularly intended to be a model which is relevant to “the challenges of developing countries” where basic services and equipment that we take for granted in wealthier countries cannot be assumed to exist.

The elements of Bing's model can be summarised as follows:

- The emphasis should not be solely on retrieval but also on “regulatory management” (“ie, preparation, drafting, adopting and reviewing statutes and regulations”).
- It should be designed “to cater simultaneously for several levels of technology”, because (at least in developing countries) telecommunications links may be unreliable and it may be necessary to rely on stand-alone workstations and paper publications.
- It should be “an integrated system which caters for the bureaucracy, the legislator, and the legal profession”. Here, as elsewhere, Bing is ambivalent about whether the system is also to cater for the needs of the ordinary public, questioning whether the same service can be used.
- A group of editors has to be established to maintain the database, which indicates a need for centralisation. Both legislation and case law are to be included in the system. The editor's tasks will include: (i) converting legal data into a standardized form (possibly including conversion from paper); (ii) “normalising” elements of that data, “especially dates and citations”; and (iii) consolidation of legislation.
- The “mother database” produced is a “production platform from which other products may be created”, including for example a legal gazette in which new legislation is made public, the (consolidated) database for retrieval purposes, printed legislation, legislation and cases on CD-ROM, or specialized collections.
- “The database will be offered online for professional users”, but not necessarily for free access.
- It “should not be a monopoly in the strictest sense”, but would for example provide publishers who wished to republish (say) statutory extracts in a textbook, the text they need in computerized form for a fee less than the cost of re-entry.
- It would provide the legislation on the Internet “preferably free of charge”. Others could hyperlink to this, which would “avoid confusion”. This free service is to meet “the lay-person's need to consult the law of the land (the principle of *publicatio legis*)”, but it is unclear whether this is to [?]

- However, the “professional” version is “a commercial system having a higher level of functionality, access to a wider range of documents and services than the free Internet solution”.
- The system therefore supports “three different user groups with needs that are not fully identical”: lay users, professional users, and regulatory managers (those developing and amending laws).
- To balance the interests of these differing users, and remove suspicion of government control of content, “a solution may be to create a *foundation* to operate the system”. It could be a separate non-profit legal entity with representatives from each of the different user groups, the courts, and academia. This would help ensure that it would “pursue the objectives of the legal system rather than purely commercial and/or political objectives”.

It is an approach to developing a national LIS which is in many respects persuasive, and which can work under the right conditions.

It found the right conditions in Norway. Jon Bing’s model sounds in many respects like a description of the very impressive Lovdata³⁵ system in Norway. Lovdata has operated since the early 1970s, and it is probably one of the few systems described in the *Handbook* which has survived and prospered to the present day.³⁶ Lovdata provides free web access to Norway’s legislation and regulations, and to Supreme Court and Appellate Court decisions, but the decisions are only available for a few months after they are handed down. Lovdata’s subscription service contains the higher court decisions permanently, plus the decisions of numerous lower courts, Norway’s treaties, law journal articles, the Norwegian version of the CELEX system for European law,³⁷ and many other valuable databases. The free access data is therefore only a small portion of the data available to paying users. One of the great successes of Lovdata is that it is self-funding from user chargers, and that funding then subsidises the limited free access provided.

The legislation as published on the web by Lovdata is now the official version of Norwegian legislation, rather than the printed version.³⁸ The system is maintained to the highest technical standards, and includes hypertext links between all of the main categories of documents.

Lovdata was established in 1981 as a private foundation by the Ministry of Justice and the Faculty of Law at the University of Oslo.³⁹ The Lovdata Board members represent the Faculty of Law at the University of Oslo, the Ministry of Justice, the Norwegian Parliament, the Norwegian Association of Judges, and the Norwegian Bar Association. Lovdata has had the good fortune to have as its Managing Director since inception Trygve Harvold, co-author with Jon Bing of their 1977 text and one of the world’s most

³⁵ The English language home page is <<http://www.lovdata.no/info/lawdata.html>>, and most of the details in the following paragraphs can be found on that page.

³⁶ I have not checked this exhaustively.

³⁷ Also offered separately by subscription.

³⁸ See Bing 2003, p 48.

³⁹ For background material leading up to the creation of Lovdata, see KS Selmer (ed), *The Lovdata Papers*, 1981, at <<http://www.lovdata.no/litt/index-papers.html>>.

experienced and distinguished contributors to the development of computerised legal research from its earliest days.

Possible problems with Jon Bing's statist model

Jon Bing's proposal has many elements of useful guidance for the development of any government-run LIS, and if successfully adopted in countries which do not have any effective access to their laws would no doubt be a big improvement. However, it is close to the antithesis of the set of principles outlined earlier as the public policies advocated by AustLII. I will go further and suggest that it is a risky model for countries that are not so fortunate as Norway for so long as it lacks this public policy element supporting full free access to law.

From Bing's model, and its Norwegian implementation, I think we can see the following potential problems.

The core problem is that there is no principle in the Bing model which holds that Courts and legislatures have a public duty to provide a copy of their output (judgments and legislation) in the best computerised form that they can produce (which changes from time to time) to anyone who wishes to publish their output, irrespective of whether that publication is for free or for fee. This is the gist of the first six of AustLII's principles.

Whether a court or legislature (or Parliamentary Counsel's Office) itself publishes its cases or legislation on the web is of secondary importance. It is often very valuable that they do so, and theirs may be the most authoritative version and even sometimes the most functional version. Yet this is of secondary importance to whether they facilitate republication by others. The same applies to Law Reform Commissions and other official sources of legal texts.

The reason for this principle being so central is the avoidance of monopolies and the facilitation of competition (or perhaps just potential competition).

I should stress that I am not referring here to an absence of Crown copyright. The theoretical right to republish legislation or cases is not worth a dime unless it is accompanied by unimpeded access to electronic copies of the source legal texts.

In Bing's model there is no requirement for either Courts/Parliament or the national LIS to provide data in electronic form to other publishers. This is clear from the fact that, even to obtain extracts of legislation, other publishers will have to pay the LIS, and that the three groups of users that the LIS is to serve do not include "other publishers".

The legislature will not even be in a position to provide other publishers with consolidated legislation, since the LIS does the consolidations of the legislation. This gives the LIS, as a revenue-generating publisher that has editorial costs as possibly its highest cost element, a vested interest in controlling access to the legislation, and limiting it as far as it can to its paying customers.

With case law, the "normalisation" of elements in the judgments is to be done by the LIS, rather than by shifting this process as far as possible back to the Courts themselves (items 1-3 in AustLII's principles).

A central body of editors at the national LIS should be avoided as far as possible, with the sources of the documents instead accepting the obligation to provide the highest quality output from their work which it is reasonable to expect, for the benefit of all potential publishers (including themselves in some cases).

Bing rightly identifies the problems of providing a service that suits both professional and lay users. However, this is in fact just the most obvious example of a deeper problem: how can any one service anticipate all the types of value-adding that might be most appreciated (even to the extent of paying for it) by a very large range of diverse audiences with quite different uses for the same core information? There are many different types of “professional” users, and an equally diverse range of lay users. The best answer to this is to encourage, or at least make possible, a multiplicity of publications.

On Bing’s model, lay users get free access only to one form of legislation (other services are to “hyperlink to it”), and no access to case law except perhaps (on the Norwegian example) to higher court decisions at the time they are likely to be featured in the newspapers or prominent in public debate. Free access becomes very second-rate access.

One missing element in Bing’s model is what happens if the services provided by the LIS are so poor that user revenues cannot pay for the service? Is it to be subsidised by government funds, and if so to what extent? I have no objection to governments paying to provide free access to legal information to the public, but this is subsidising privileged access to the legal profession, which is different. The main problem is that there is no limit to the extent of subsidy to the LIS (and therefore no efficiency requirement) – in short, no exit strategy. In the worst-case scenario, it is a recipe for increasing subsidies to a body providing a second-rate service but shielded from competition.

The idea of a non-profit foundation to run a LIS is a good one, and is similar to the BAILII Trust that runs the UK LII. But a non-profit monopoly is still a monopoly. The history of legal information retrieval is littered with failed non-profit government systems, including those that denied others access to the data they needed.⁴⁰ Bing’s model does not require involvement of academic bodies, but just encourages this. However, Lovdata was jointly initiated by Oslo University, and so is one of the earliest examples of the “university-based” LIS that have become a feature of the free-access-to-law movement. This may be a more important element in Lovdata’s success than Bing’s model concedes.

Norway has clearly avoided such problems, so they are not inevitable. But how many systems will have the luck to be led by a Trygve Harvold since inception? How many countries have both the technical sophistication and the lack of corruption of Norway? How many will make the choice to have the system “university-based” from the start?

If we need a convenient label, I would describe this as a “statist” model of a national legal information service, where a government-created body (a “quango” in fact)⁴¹ is given a de facto monopoly over the provision of legal information, with only very limited (if any) obligations to provide free access. The quantity and quality of free access becomes the gift of the quango. It may turn out to be substantial, as in Norway, but that is just good fortune and may not be repeated elsewhere very often.

⁴⁰ The UK government’s project for computerised consolidated legislation, now nearing its 15th birthday, still refuses to provide anyone with whatever consolidated UK legislation it has produced. Australia’s dual monopolies retarded the development of any LIS for over a decade: see further G Greenleaf, A Mowbray and G King, “New directions in law via the internet – The AustLII Papers”, *Journal of Information, Law and Technology* (JILT), 1997, issue 2, at <http://elj.warwick.ac.uk/jilt/issue/1997_2>.

⁴¹ That is, “quasi-autonomous non-government organisation”.

An alternative competitive model based on full free access

An alternative theoretical model, which is also consistent with the history of the free-access-to-law movement and the Legal Information Institutes it has produced, places the emphasis on something resembling the set of seven “AustLII principles” supporting access to law. This emphasis is shared by other LIIs, as it is implied by the Declaration on Free Access to Law. It is also illustrated by the comment of the founder of CanLII, LexUM’s Daniel Poulin, that “open access”

... refers to a more open-ended form of publication. In order to be truly open, publication must not be restricted by any conditions as to subsequent use or republication of the documents, except for conditions to ensure accuracy.⁴²

“Full free access” is therefore not based so much on the public’s free access to any particular legal information system (that is only a possible by-product) but rather on the obligation of public bodies to provide free access to the computerised sources of legal data to those who wish to publish them.

Competition then provides most of the rest of what is required for high quality public access to law, including free access. Bing’s model, suitably adjusted, can then be a recipe for the effective development of a national LIS free from worries about monopoly abuses.

According to this model, consolidations of legislation are not done by the LII, but by the offices of Parliamentary Counsel (or equivalent) in each jurisdiction, which supply them free of charge to other publishers including the LIIs. All nine Australian jurisdictions now manage to do this, and so do most jurisdictions in Canada, so it is no longer “rocket science”.⁴³

With case law, AustLII receives data from around 100 Courts and Tribunals, but does not edit the decisions of any of them. Most of its case law efforts over the past decade have gone into assisting the Courts and Tribunals to improve the quality and consistency of the data they supply to AustLII and other publishers, and (most importantly) to accept the responsibility for this quality.

The LIIs in different countries have all emerged with different models of funding, none based on direct user charges, but variously drawing on collective contributions from their users (lawyers, business groups etc), or from the bodies whose data they publish (agencies, courts etc), or from academic grants, or from aid organisations or government grants. It does not matter too much, because they all have to compete (on this model) against any other publishers (for fee or for free) who wish to attempt to provide a more attractive service using similar data.

To date, the self-conscious participants in the free-access-to-law movement have been university-based LIIs, and while this does seem a good formula (with similar virtues to the foundation Bing proposes and consistent with it), LIIs based on government agencies (as LegiFrance may become) or civil society bodies, may make just as valuable a contribution.

Of course, AustLII has its own model as to how a successful LII can be established in developed or developing countries, and LexUM has its own model which it is successfully pursuing.⁴⁴ Yet they are not the point of this article, or, more importantly, the essence of the “competitive” model for developing a national LIS. This model is not so

⁴² Poulin, *op cit*.

⁴³ Even if they cannot yet manage it in the UK.

⁴⁴ See Poulin, *op cit* for a detailed articulation.

much a recipe for a national LIS as a set of pre-conditions for how different recipes (such as Bing's, or AustLII's, or LexUM's) may safely be followed.

Language and law – A Nordic LII?

LexUM's Daniel Poulin has outlined a challenging problem⁴⁵ for those outside English-speaking countries who are involved with legal information systems:

To date, Anglo-Saxon countries have been more pro-active with regards to the dissemination of their law. In comparison, Francophones, Hispanophones and most jurists from other linguistic groups have remained rather passive in this regard. Radically amending this situation, therefore, is of the utmost importance to offer the non-English speaking world better access to foreign law.

The prevalence of the law of Anglo-Saxon countries on the Web also means that the common law tradition has been made much more accessible than other traditions. Should this dominance prevail, the manner in which statutes are drafted, the construction of legal arguments and the overall spirit of the common law will become the only model available for purposes of consultation for those who want to compare and improve their own legal systems.

Hence, to avoid a unilingual, single tradition legal context, a better dissemination of the law from countries with civil law traditions should be promoted. Better access to law from francophone countries could create a shared network where these similar legal systems can mutually influence each other. The same ideals can be applied to other linguistic spheres and legal traditions, which to date have rarely been distributed or accessed.

Furthermore, if the open access to law national sites are complemented by a regional or, in broader terms, by a civil law portal, more benefits are gained. Gathering and archiving legal texts from similar legal traditions, similar national legal systems or from a shared language obviously enhances open access. Such resources would constitute an indispensable site for jurists working in languages other than English by increasing the possibilities to study the multiplicity of existing legal systems.

Poulin and his LexUM colleagues have taken a large step toward redressing this problem in the francophone world with the launch of Droit Francophone. Further, although not an explicitly language-based approach, regional free access law portals have started to emerge where there is a distinctive shared legal culture. This is most notable in the case of the Pacific Islands Legal Information Institute (PacLII)⁴⁶ to serve as a portal for legal systems with Melanesian and Polynesian cultural content, and in particular to facilitate legal education and comparative jurisprudence across all fourteen countries involved.

With Jon Bing as one of the central figures, the Nordic countries as a whole led the world from the 1970s onward in research into the global development of legal information systems, and in terms of practical developments few systems could compare with Lovdata.

It is therefore a little surprising that there has not yet developed any overall Nordic or Scandinavian portal for legal information for a group of countries that have many linguistic similarities and a distinctive and significant legal tradition. So this paper finishes with some questions I cannot answer. Would the legal cultures of the Nordic countries or their Scandinavian subset be advanced by them sharing some common free

⁴⁵ Poulin, *op cit*, section entitled "Diverse Traditions and Different Languages".

⁴⁶ See <<http://www.pacii.org>>.

access service? It might be easier for the Scandinavian countries (as opposed to a broader Nordic grouping) to achieve such a service given their closer linguistic ties. Would the rest of the world benefit, in the ways Poulin suggests, from inclusion of Nordic or Scandinavian content in the emerging global networks? Would Europe benefit from Nordic or Scandinavian leadership in the free-access-to-law movement?

Antipodeans

A former Dean of the Law Faculty of University of New South Wales, when welcoming a delegation from the Norwegian Research Centre for Computers and the Law (NRCCL), commented that in the field of human rights Norway was renowned for “playing above its weight”,⁴⁷ and it struck me at the time that this was an apt description of Norway’s global role in the fields of cyberspace law and computerisation of law. With the NRCCL and Lovdata as the two central institutions, and Jon Bing, Knut Selmer, Olav Torvund and Trygve Harvold among these institutions’ leading individuals, Norway’s reputation is well above its size.

In another Antipodes, Australians involved in these areas of research have always been conscious of Norway’s leading role, and since the 1970s have regarded the NRCCL and its work as an institution to emulate (we haven’t managed to). Jon Bing’s contributions to the analysis and development of legal information retrieval systems, and to the broader field of computerisation of legal information, have deservedly played a major part in establishing Norway’s reputation in these fields.

⁴⁷ A sporting reference to a small player who is well able to hold his/her own in the “big league”.