68th IFLA Council and General Conference
August 18-24, 2002

Code Number: **116-163-E**
Division Number: **I**
Professional Group: National Libraries
Joint Meeting with: Information Technology
Meeting Number: **163**
Simultaneous Interpretation: **-**

Archiving the Web – some legal aspects

Alenka Kavčič-Čolić
National and University Library
Ljubljana, Slovenia,
E-mail: alenka.kavcic@nuk.uni-lj.si

Abstract:

Technological development has changed the concepts of a publication, reproduction and distribution. The legislation, together with the Legal Deposit Law does not incorporate these changes, and is very restrictive in the sense of protecting the rights of authors of all electronic publications. National libraries and national archival institutions, being aware of their important role in preserving of the written and spoken cultural heritage, try to find different legal ways to its realization. The paper presents some legal aspects of archiving the web pages, concerning the harvesting, providing public access to them, and long-term preservation.

Keywords: world wide web, web pages, electronic publications, legal aspects, legal deposit, national libraries

1. INTRODUCTION

The report of the Committee on Intellectual Property Rights and the Emerging Information Infrastructure, the Computer Science & Telecommunications Board, and the Commission on Physical Sciences,

Mathematics, and Aplicatons: *The Digital Dilemma: Intellectual Property in the Information Age* (1999) lists three characteristics of the technological development which had effect on drastic changes in economics of information: (1) information in digital form has changed the economics and character of **reproduction**, (2) computer networks have changed the economics of **information distribution**, and (3) the World Wide Web has changed the economics of **publication** [Samuelson & Davis, 2000].

The digital nature of electronic publication has changed the concept of **reproduction**. Never before was possible to cheaper and faster reproduce a publication without the loss of quality. Electronic copies are identical to the original.

Digital networks have brought an end to the traditional meaning of **information distribution**, which had to do only with *tangible materials*. The control of booksellers over the destiny of any material ended at the point of receiving the payment or the subscription fee. In the physical sense, the new owner could do whatever he wished with the book he had bought, sell, loan, or even destroye it. In the legal practice this is known as the *first-sale rule*. It is no longer so with electronic publications. The possibilities of their reproduction are not limited, therefore for the purpose of accessing certain material there has to be established a continuing bond between the author or the provider and the end user in the form of a contract or a licence in order to avoid breaking the copyright and endangering the commercial interests of the carriers of copyright. Such contracts of course are time limited, i.e. after their validity ceases the user loses all rights of accessing the contracted material.

The development of the internet has also changed the meaning of a **publication**. Today, anybody can be author or publisher of an electronic publication. The statistics show that the average life expectancy of a web page moves between 44 days and 2 years, and that very few of those have changed their contents in one year [Kenney … et al., 2002]. Such vast creativity causes difficulties in controlling the rights of individual authors.

We are aware that the legislation has changed slowly, by all means slowlier than the information technology. Most legislative systems have not foreseen the technological development in the direction of electronic publications and still use the traditional concepts of a publication and its reproduction and distribution, which are useful only in case of tangible publications. The best proof of that is the legal deposit legislation.

2. LEGAL DEPOSIT LAW

While the form and carriers of a publication have changed over the centuries, the basic functions of national libraries have remained almost unchanged. During the last hundred years their most important task has been the preservation of written and spoken cultural heritage. This function is in most countries supported by the Legal Deposit Law. The content of this law, however, varies from one country to another.

In most European and some other countries (Canada, USA, Australia) the Legal Deposit Law covers mainly printed materials and electronic documents on physical carriers. The exceptions to this rule are Norway and Denmark, where the Legal Deposit Law also includes the networked publications, and Switzerland and the Netherlands which have no such law at all [Martin, 1999].

Every national library has its own way of fighting to obtain the national collection. The co-operation with the publishers and publishers' unions is becoming a more and more important approach in this respect. In the Netherlands, for example, there does not exist a legal deposit law, however, they do have an agreement with the Publishers' Union which enables them to receive legal deposit. Another example are the project ELEKTRA and EVA in Finland. Such examples show that the national libraries have taken upon

themselves to respect the conditions set by the publishers, e.g. access moratoria, controlled or limited access, etc.

Within the CoBRA+ programme a Joint Committee was set up to represent the CENL (Conference of European National Librarians) and the FEP (Federation of European Publishers). The task of this commission was to find a form of co-operation, which would be acceptable for both, the publishers and the national libraries. The result of the negotiations is the *Code of practice for the voluntary deposit of electronic publications.*

The contracts with the publishers can regulate the access to individual documents, however, they are not a solution in the case of capturing the entire web. In Australia, for example, within the framework of the project PANDORA, an excellent methodology (including the forms and accurate instructions) for a selective processing of electronic publications from the web was developed. However, the legal and administrative processes are complex, and their finding was that the web capturing is five times as expensive than the purchase of the printed materials [Bergamin, 2000]. The number of electronic publication is growing much faster than the number of printed ones. In Europe there can be found some examples of a non-selective approach, i.e. capturing the entire web. We would like to mention Kulturarw3 in Sweden, and Finland where a web harvester has been developed in the project NEDLIB. In the USA a non-profit company Internet Archive has been systematically collecting the entire web since 1996. [Kavčič-Čolić, 2001]

Earlier practice has shown that the national libraries and archival institutions cannot wait to see the legal deposit law change and feel it is their responsibility and duty to take certain actions towards the preservation of the present for the future. While doing this they should respect the current legislation in the field of copyright.

3. ARCHIVING THE WEB

Archiving web pages requires attention on three aspects which have different legal basis: (1) the procedure of harvesting the web pages and electronic documents on the internet; (2) enabling the public to access them; and (3) their preservation for the future.

3.1 Harvesting the web

In the countries where legal deposit legislation does not cover the intangible electronic publication the building of the national collection is regulated by the copyright legislation.

What in fact means copyright? In the USA the copyright is "the limited monopoly created for the purpose of providing people with a financial incentive to create copyrightable materials, to create works of literature, art, …", in Europe it means "the right of every author to control the reproduction of the products of his or her own brain […and] is seen […] as an extension of the author's personality" [Strong, 1994]. Both statements are two different sides of the same concept of copyright protection.

The foundations of the current copyright protection have been set at the international conference in Bern in 1886. The last revision of this convention happened in Paris in 1997. The most important article concerning the copyright is 9(2) which states that "the national legislatures may authorise the reproduction of copyright works in 'certain special cases, provided that such reproduction does not conflict with a normal exploitation of the work and does not unreasonably prejudice the legitimate interests of the author' [Wall, 1998, p.338]". In 1967 the World Intellectual Property Organisation (WIPO) was established which took responsibility for administering of many international conventions and agreements on intellectual property and copyright, among them also the Bern Convention. The most important WIPO conference took place in Geneve in December 1996. It focused on possible changes of the Bern Convention. They

proposed, among others, the right of the users to "browsing, concerning viewing of screened material without permission" [Wall, 1998, p.339]. Naturally, the publishers' lobbies were stronger, and this right has not been given consideration even in the Directive 96/9/EC of the European Union on the legal protection of databases nor in the Directive 2001/29/EC on the harmonisation of certain aspects of copyright and related rights. In the latter, the following rights of authors have been stated:

- "[…] **right to authorise or prohibit direct or indirect, temporary or permanent reproduction** by any means and in any form, **in whole** or in part;
- […] **right to authorise or prohibit any communication to the public** of their works, **by wire or wireless** means, including the making available to the public of their works in such a way that **members of the public may access them from a place and at a time individually chosen by them**;
- […] **the exclusive right to authorise or prohibit any form of distribution to the public by sale or otherwise** […]" [emphasis added].

The 5th Article of the Directive states the exceptions which are "in the public interest for the purpose of education and teaching", specifically "for private use and for ends that are neither directly nor indirectly commercial, on condition that the rightholders receive fair compensation". The reproduction is limited to parts and does not cover whole documents.

Within the EU legislation the works under copyright are computer programmes, interfaces, databases and all author creations regardless of their carriers. Web pages as such are not mentioned in these directives. As a group of files which form a unity, they belong among author works and can therefore be included into the Copyright law. The Copyright Licencing Agency in the Great Britain states that "The World Wide Web is subject to copyright, and Web pages are themselves literary works"[1]. Any collecting or archiving of these materials, without the permission of the author or copyright holder is against the law. Exceptions are individual data, government publications which belong into the public domain and other publication in which it is explicitly stated that their reproduction is allowed. Any further use of such publications requires the citing of the source. The rights of authors are limited to their life plus 70 years after the death of the author.

Since the operation of many applications means that the reading the electronic documents or accessing the databases requires automatic reproduction of certain elements or even the whole application to the network or to a computer, the Directive 96/9/EC, which regards the computer software and databases allows the temporary reproduction as exception, in the sense of caching, hosting, and browsing.

In the case of the electronic publications where the access has been enabled through a licence, there could be negotiated with the publisher an agreement on the transfer of the whole collection with the goal of preservation of the material.

3.2 Enabling the public to access the electronic documents including web pages

On the basis of the Article 19 of the United Nations' Universal Declaration of Human rights "Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to *seek, receive and impart information and ideas through any media and regardless of frontiers*" [italics added].[2]

We would like to mention two more statements by the IFLA Committee on Free Access to Information and Freedom of Expression (FAIFE)[3]:

[1] URL: http://www.cla.co.uk/copyrightvillage/internet.html
[2] URL: http://www.faife.dk/
[3] Ibid.

- "Libraries provide access to information, ideas and works of imagination. They serve as gateways to knowledge, thought and culture."
- "Libraries have a responsibility both to guarantee and to facilitate access to expressions of knowledge and intellectual activity. To this end, libraries shall acquire, preserve and make available the widest variety of materials, reflecting the plurality and diversity of society."

The European Copyright User Platform (ECUP) has also adopted a Position Paper which "purpose was to outline and justify the lawful uses of copyrighted works by individuals and libraries in the electronic envirnoment" [ECUP, 1997, p.386].

But the European legislation does not support free access to materials without the permission of copyright holders. The Directive 2001/29/EC as well as the Directive 1996/9/EC both state the exclusive right of authors the right to make the work public, including the right to give the public access to it.

From our point of view it is important the article 3 of the Directive 2001/29/EC, which defines "the right to authorise or prohibit any communication to the public of their works, by wire or wireless means, including the making available to the public of their works in such a way that **members of the public may access them from a place and at a time individually chosen by them"** [emphasis added]**.** This second part protects any form of communication with the public or allowing the access to the members of the public which are not present at the place from which the publication or the access derive [Šetinc, 2001].

The same Directive in the Article 5 allows "specific acts of reproduction made by publicly accessible libraries, educational establishments or museums, or by archives, which are not for direct or indirect economic or commercial advantage". This exception is in agreement with the the doctrine of *fair dealing* characteristic for the anglo-saxon legislations (*fair use* in the USA), which "is generally thought to refer to restricting the amount of material which can be copied or used fairly so as not to damage the economic interests of rights owners" [Pedley, 1998, p.29]. The goal of the *fair dealing* is in fact to be "a balancing point between the rights of authors to exclusive control over their work and the rights of users to have free access to the ideas contained in those works" [Harper, 2001].

If we take into account the 19th Article of the United Nations' Universal Declaration of Human Rights, the mission of the libraries in providing acces to information and the *fair dealing* doctrine, then certainly the libraries should have not only the right but also a duty to give access to electronic material. Maybe national libraries and other archival institutions could achieve a compromise, should they manage to find adequate technological solutions to help them to make the copyrighted materials available to the public for on-the-spot reference use [Šetinc, 2001]. Similarly as with the physical materials the users could in the national library have the right to view the electronic materials without the right to copy it. This would enable the protection of commercial and other interests of the authors.

Harvesting the web does not mean giving the public the right to view it. Unless if it is explicitly stated in the web page or if a certain amount of time has passed, the protection of the copyright has ceased. The characteristic of the electronic publications however, is rapid growth and ephemerity. It is very likely that after some time none of the electronic documents will survive, mainly because of the technological and software development.

3.3 Preservation of the web pages for the future

For some years scientists have been researching the effects of time and technological development on the existence of the electronic publications. The project NEDLIB has dedicated special attention to this issue. The today's solutions for the long-term preservation are mainly: the museum approach (preservation of the technological and software environment to access the electronic publication), emulation (access to the

document with an additional interface), and migration (the conversion of the application of the electronic publication to a higher version or a more contemporary application). Certainly a solution has to be found which concernes the existence of the carriers themselves. The migration and emulation require changes in the form of the document. In these cases it means violation of the copyright. Concerning the fact that the changes would have to be frequent, the document would need several interventions. In case this was left to the publishers and the authors, the question arises whether they would be interested to perform it. Many of them follow the needs of customers, which are focused to the latest literature. Therefore there is a great danger that a large part of this would be thrown away forever.

For this reason it is very important that the national libraries and national archival institutions collect and accept the responsibility for these materials. Very few, however, have a legislative background for such actions without the permission of the authors or the publishers.

4. CONCLUSION

The arrangement of the copyright individually, for every web page and with every author would require from the national libraries and other archival institutions a strict policy on selection, a lot of time, and human and financial resources. It involves the risk of losing a major part of important electronic documents on the web. Similarly as with the materials obtained throught the legal deposit, the national libraries should collect everything that has been published and which represents our cultural, historical and scientific image, regardless of the carrier. In the same way, these materials should be accessible to the public, even if only for reference use. Another important task of national libraries and archives is the long-term preservation of these materials, inspite the fact that this means a direct intervention in the document. The question is if this is allowed without the permission of the author or the publisher? According to the valid legislation, no. However, it is in the interest of the nation and mankind, which should urge the governments of all countries to uncompromisingly and as soon as possible accept such actions within the legal deposit legislation.

Most governments protect the commercial interests of the publishers and the authors. But they do not take into account the market itself which has already begun its own way of protecting itself with various technological means [Schlachter, 1997]. For this reason the national libraries and national archival institutions should receive greater government ald legislative support in their striving to preserve the whole written and spoken heritage on all carriers. This would enable these important institutions to uninterruptedly perform their duties in preserving the present for the future generations.

5. REFERENCES

Bergamin, Giovanni (2000): "A standard for the legal deposit of on-line publications". In: Connolly, Pauline A. (ed.): *The digital library : challenges and solutions for the new millennium* : proceedings of an International conference held in Bologna, Italy, June 1999. - Boston Spa : IFLA, Offices for UAP and International Lending, pp. 119-127.

CENL & FEP (1999): *Code of practice for the voluntary deposit of electronic publications***.** URL: http://www.bl.uk/gabriel/fep (4-29-2002).

Committee on Intellectual Property Rights and the Emerging Information Infrastructure, the Computer Science & Telecommunications Board, and the Commission on Physical Sciences, Mathematics, and Aplicatons (1999): *The Digital Dilemma: Intellectual Property in the Information Age.* - Washington : National Academy Press. URL: http://bob.nap.edu/html/digital_dilemma/ (4-29-2002).

Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society. In: *Official Journal* L 167, 22/06/2001, P.0010-0019.

Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases. In:*Official Journal* L 077 , 27/03/1996 P. 0020 - 0028

ECUP (1997): "Position on user rights in electronic publications". In: *The Liber Quarterly*, No.7, pp. 386-395.

Harper, Georgia (2001): *Will we need fair use in the twenty-first century?* URL: http://www.utsystem.edu/ogc/intellectualproperty/fair_use.htm (4-18-2002)

Kavčič-Čolić, Alenka (2001): "Elektronsko arhiviranje spletnih strani: nov izziv za nacionalne knjižnice". In: *Digitalna knjižnica*: [zbornik referatov] / ZBDS, strokovno posvetovanje, Radenci, 10-12 October 2001 : [Melita Ambrožič Chief Editor]. - Ljubljana : ZBDS, pp. 211-224.

Kenney, Anne R. … et al. (2002): "Preservation risk management for web resources: virtual remote control in Cornell's project prism". In: *D-Lib Magazine*, 8(1). URL: http://mirrored.ukoln.ac.uk/lis-journals/dlib/dlib/dlib/january02/kenney/01kenney.html (1-2-2002).

Martin, Elizabeth (1999): Management of networked electronic publications : a table of status in various countries. - Ottawa : National Library of Canada.

Pedley, Paul (1998): *Copyright for library and information service*. - London : Aslib.

Samuelson, Pamela and Randall Davis (2000): *The digital dilemma: a perspective on intellectual property in the information age*, paper written for presentation at the 28th Annual Telecommunications Policy research Conference. URL: http://www.sims.berkeley.edu/~pam/papers/digdilsyn.pdf (5-13-02), p.7.

Schlachter, Eric (1997): "The intellectual property renaissance in cyberspace: why copyright law could be unimportant on the internet". In: *12 Berkeley Technology Law Journal* 15-51 [course packet]. URL: http://www.law.berkeley.edu/journals/btlj/articles/12_1/Schlachter/html/text.html (29-4-2002).

Strong, William S. (1994): *Copyright in the new world of electronic publishing*. Presented at the workshop*Electronic Publishing Issues II* at the Association of American University Presses (AAUP) Annual Meeting, June 17, 1994, Washington, D.C.. URL: http://www.press.umich.edu/jep/works/strong.copyright.html (4-29-2002)

Šetinc, Lenart (2001): "Nekatera avtorskopravna vprašanja digitalne knjižnice". In: *Digitalna knjižnica :* [zbornik referatov] / ZBDS, strokovno posvetovanje, Radenci, 10-12 October 2001 : [Melita Ambrožič Chief Editor]. - Ljubljana : ZBDS. - pp. 60-71.

Wall, Raymond A. (1998): *Copyright made easier*. - 2nd ed.. - London : Aslib.