Annex 2

WEB HARVESTING, Case Study Finland by Tarja Koskinen-Olsson

1. Legislative Framework

Legal deposit legislation was recently renewed in Finland and the new statute[1] entered into force on 1 January 2008.

The legislation is generic and covers the legal deposit of different genres of cultural heritage. It replaces previous separate statutes providing for the deposit of print material and films. It now also includes national television and radio programs and web-based material. The legislation has references to copyright legislation.

In implementing the Information Society Directive, Finland included the relevant provisions concerning deposit copies in its copyright law, with the date of coming into force of some provisions to be prescribed to coincide with the entry into force of the legal deposit legislation.

2. Legal Deposit Legislation

The need to renew the legal deposit legislation arose from the new media and communication environment and the development of the information society. The aim was to cover long term preservation of all cultural material to serve the needs of researchers and others.

This outline concentrates on web-based material, referred to below as internet resources[2]. The law on legal deposit and preservation of cultural material includes the following provisions[3] concerning internet resources:

- Scope of application (2 §): resources available from servers situated in Finland and other internet resources that are meant for the public in Finland;
- Definition (3 §): internet resources are materials available in information networks.

Chapter 3 of the law concerns internet resources and the following is a general summary of the provisions:

- Web-harvesting and deposit: The National Library shall gather/harvest and deposit internet resources from the net, and the material shall make a representative and diversified picture/sample of all material that is made available for the public over the net at different times;
- The web publisher shall enable the harvesting of material or deposit the material himself, if harvesting through technical means is not possible (the web publisher thus has a secondary liability for deposit);

---

[1] Act on Legal Deposit and Preservation of Cultural Material (1433/2007).
[2] Information is based on discussions with Mr. Juha Hakala, Director of IT Development, the National Library of Finland.
[3] Unofficial translations.

- Special provisions apply to material that cannot be deposited due to a technical reason;
- The Ministry of Education and Culture shall confirm the plan of the Finnish National Library concerning the volume and frequency of harvesting and the conditions of delivery; the internet resources shall be deposited in such a way that the original time and place of the resources are recorded.

The National Library has in practice harvested material from the net since 2000. Extensive harvesting takes place once a year; moreover special collections on selected topics are gathered. More than 50 million resources are harvested annually.

Within the framework of IIPC (International Internet Preservation Consortium) technical tools for web-harvesting have been developed and are currently used by the Internet Archive[4] and several national libraries, including the National Library of Finland, under the name "Heritrix". These open source applications are freely available at http://crawler.archive.org/.

The aim of the Finnish National Library is to harvest the following internet resources:

- All internet resources made available from Finnish sites, regardless of the domain name (it can be fi, com, net, org or anything else);
- All Finnish language sites anywhere on the net;
- All sites that include information about Finland.

Harvesting is based on exhaustive list of Finnish servers. The systematic gathering uses more than 50.000 root pages as starting points.

3. Copyright Legislation

The implementation of the Information Society Directive brought about amendments to the Finnish Copyright Act, which entered into force on 1 January 2006. Some legal deposit provisions were activated with the passage of the legal deposit legislation and entered into force on 1 January 2008.

The provision concerning web-harvesting in the Finnish Copyright Act (section 16b) is entitled "Use of works in legal deposit libraries". The relevant provision, section 16b (1), paragraph 3, reads as follows:

Make copies of works made available to the public in an open information network for inclusion in its collections.[5]

The on-site use of the works in public libraries is made possible by Article 5. 3 (n) of the Information Society Directive:

[4] The Internet Archive (US) is a private foundation that harvests web resources from all over the world, including Finnish websites. It is estimated that the collections include some 60 billion pages. Some libraries have bought older information concerning their countries from the Internet Archive foundation.
[5] Unofficial translation.

An archive or a library open to the public, to be determined in a Government Degree, may, unless the purpose is to produce direct or indirect financial gain, communicate a work made public that it has in its collections, to a member of the public for purposes of research or private study on a device reserved for communication to the public on the premises of the institution. This shall be subject to the provisions that the communication can take place without prejudice to the purchasing, licensing and other terms covering the use of the work and that digital reproduction of the work other than reproduction required of the use referred to in this subsection is prevented, and provided that the further communication of the work has been prevented (Section 16a (2)).

Thus the copyright legislation made provision for the needs of legal deposit and enabled web-harvesting.

4. Evaluation

The provisions in the legal deposit legislation covering cultural material and stipulations in copyright legislation are well suited to cover the needs of web-harvesting.