

Law and the Would-be Web Archivist

Andrew Charlesworth

Director, Centre for IT and Law

University of Bristol, UK

Introduction

- On a personal note – the frustration & futility of legal academic citation to web resources:
 - One book chapter - 65% of web documents cited to removed or URLs obsolete within a year.
 - One journal article - 50% of web documents cited to removed or URLs obsolete before publication.
- Average lifetime of a web-based document has been estimated at approximately 44 days.
- Valuable web materials may be irretrievably lost.

Preservation of Digital Materials

- Digital materials would seem relatively simple to preserve:
 - easily copied with little or no loss of quality;
 - stored in compact digital formats.
- Yet digital materials have their problems too:
 - decaying or obsolete media;
 - proprietary compression or DRM technologies.
- But the key obstacle may often be the laws applied to the technology or to the material.

Preservation of Web Materials

- JISC/Wellcome Trust Web Archiving Study
 - Examined the law in UK, EU, US, Australia.
- Broad Overview of Legal Issues of Concern to Web Archivists
 - National laws on legal deposit of digital materials;
 - Intellectual Property laws, esp. copyright;
 - Defamation law, esp. libel law;
 - Privacy law, esp. data protection law;
 - Content liability laws, esp. obscenity/indecency laws.

Legal Deposit

- National approaches to the legal deposit of digital materials vary widely.
 - UK - does not apply to publication in microfiche, CD-ROMs, databases or any new non-book media.
 - US - deposits of networked electronic publications solely available online are not presently required.
 - Denmark - for an Internet document to be subject to legal deposit it must be 'static' not 'dynamic'.
 - Sweden - the Royal Library can collect material from Swedish web sites on the Internet and to allow public access to it within library premises.

Intellectual Property

- In most jurisdictions, if no statutory right to archive webpages pertains, permission to copy the works must be obtained from rightholders.
 - UK – libraries/archives are allowed to make a copy from any item in their permanent collection for preservation & replacement.
- Determining who rightholders are may simply be impossible for large scale web archiving operations in the absence of © metadata in HTML etc.

Defamation

- Defamation law varies widely between nations.
- In most (but not all) jurisdictions,
 - the basis of defamation liability is the publication of untrue information; and,
 - liability is based on the extent of the damage to the reputation of the person referred to; and,
 - a person's reputation cannot be damaged unless the information is disseminated to other people.
- Often republication (including in archival form) gives rise to further causes of action.

Privacy

- A web archive will almost inevitably contain personal data, e.g. names, addresses, work and home telephone numbers.
- Collection and storage of such data in an archive may result in personal data becoming, and remaining, available for search and retrieval.
- The larger the archive, the more difficult it will be to ensure DP laws are observed, particularly if the archive is made publicly available.
- Archiving exemptions in laws such as the UK Data Protection Act do not fit comfortably with the archiving of public webpages.

Content liability

- Content liability law shows little uniformity in the EU:
 - Some Member States operate rigorous regimes of censorship over depictions of sexual activity, e.g. the UK
 - Others take a rather more *laissez faire* approach to their citizens' proclivities, e.g, the Netherlands
 - Some Member States bar certain other types of material e.g. neo-Nazi or holocaust denial material
- Some aspects of content liability show a degree of consensus e.g. the undesirability of child pornography,
- BUT this consensus does not appear to extend to the uniform interpretation of subject matter, uniform definition of offences, or uniformity of punishment.

Solutions: Archive and Hope

- A web archive might simply archive 'public' webpages & try to avoid legal complications by:
 - Careful publicity designed to show the value of a web archive for the public and future researchers
 - Mechanisms to prevent the collection of certain © materials, or to later withdraw them from the archive
 - Mechanisms to allow for the removal of material which breaches libel, DP or other laws
 - Having a time lag between the collection of the material and its appearance in the archive.
- Experience suggests this may not be enough.

Solutions: The Closed Web Archive

- An archive 'sealed' to public access for a set period runs a reduced legal risk over its content:
 - it won't affect ©holders' economic rights in works;
 - sealing the information reduces or removes the risk of defamation and content liability.
- Yet a web archive that denies public access:
 - loses much of its utility in the short to medium term;
 - will struggle to raise awareness of its existence;
 - will struggle to attract public or private funding.

Solutions: National Legislation

- National legislation may:
 - explicitly provide permission to make copies of digital works for preservation and archiving;
 - provide for the legal deposit of digital works;
 - prevent prosecution for criminal offences, or civil suits merely for archiving certain types of work.
- But:
 - national legislatures move very slowly;
 - digital archiving is not a ‘hot’ political topic;
 - there may be significant counter-lobbying.

Solutions: International Action

- International consensus on the need to provide legal backing for archiving of web materials would be valuable.
- But existing international organisations are often of little help because:
 - their decisions/resolutions etc. carry no weight; or
 - they are dominated by special interest groups whose agenda does not include preservation and archiving, or is concerned solely with maintenance of commercial advantage.

Solutions: Selective Archiving

- A web archive may negotiate with a limited no. of webpage owners for the right to archive:
 - reduces the amount of material to be surveyed for legal problems;
 - may reduce the risk of legal liability by shifting some of the burden to the webpage owners;
 - may allow for the archiving of 'deep Web' resources by agreement with publishers and rightholders
 - BUT may struggle to with nontraditional publishers, and fail to collect representative web materials.

Conclusions I

- Web archivists in the EU and elsewhere face a complex and confused set of legal rules.
- When archiving the ‘national public web’ it may be hard to comply fully with legal obligations.
- Archiving the ‘international public web’ can only be done on an ‘archive and hope’ basis.
- Archiving the ‘deep web’ may be possible on a national basis with static materials under national legal deposit rules, but otherwise may have to be carried out under licence.

Conclusions II

- Various web archiving models already exist
 - National Library of Australia's PANDORA project.
 - National Library of France project.
 - Royal Library of Sweden's Kulturarw³ project.
 - Royal Library of Denmark's Netarchive.dk project
 - The US Internet Archive.
- These offer different legal risk management strategies, varying degrees of legal risk, and varying levels of web archiving.