Digital Preservation **Handbook**

Content Specific Preservation

Illustrations by Jørgen Stamp digitalbevaring.dk CC BY 2.5 Denmark

Who is it for?

Operational managers (DigCurV Manager Lens) and staff (DigCurV Practitioner Lens) in repositories, publishers and other data creators, third party service providers.

Assumed level of knowledge

Novice to Intermediate

Purpose

- To provide a bridge to and achieve synergies with, reports in the DPC Tech Watch Series. The reports provide advanced level "deep dives" in specific areas of content preservation (e.g. email) that can be cited or to source case studies in the Handbook.
- To be developed for ease of maintenance, cost-efficiency, and sustainability in the long-term by the DPC via updates and additions to the Tech Watch series.
- To provide a brief overview and case studies, suitable for novice or intermediate level users, of digital preservation issues for specific content types covered by DPC Technology Watch Reports. Currently three content types are available: e-journals, moving picture and sound, and web-archiving. We hope to add more at a later date.

Reusing this information

Contents

e-Journals

Illustration by Jørgen Stamp digitalbevaring.dk CC BY 2.5 Denmark

Overview

This case study provides a brief novice to intermediate level overview for e-journal preservation summarised from the DPC Technology Watch Report Preservation, Trust and Continuing Access for e-Journals with updates and additions by the author. Two "mini case studies" are included together with short summaries of major services and solutions. The report itself is recommended to readers who need a more advanced level briefing on the topic and practice. It covers a wider range of issues and practice in greater depth with extensive further reading and advice (Beagrie, 2013).

Introduction

Digital Preservation and trust in having continuing future access to digital content have become increasingly important for research libraries as published journals and articles have shifted from print to electronic formats. Traditional publishing business models and relationships have also undergone major transformations as a result of that shift.

Among many significant changes there has been a move from libraries purchasing and physically holding (and preserving) a paper journal locally (with multiple redundancy of copies between libraries), to renting (licensing) remote access to an electronic journal held on publishers' platforms that are often based internationally in other jurisdictions.

In parallel, there has been a growing open-access movement for e-journal articles that seeks to remove the subscription charges for access. Subscription journals, open-access journals and hybrids of the two (either a mixture of open-access and subscription articles in a journal or a 'moving wall' to open access after a fixed period of time) provide a complex landscape for the preservation of, and long-term access to, e-journals.

This e-journal landscape continues to evolve as e-publishing itself begins to shift from static to dynamic content, and the importance of data and supplementary material linked to articles increases in major disciplines.

All these changes in turn have made preservation of e-journals more demanding, more international and dependent on others, and brought issues of trust to the fore. Trust in this context is not solely of technology for preservation, but negotiating rights (and retaining a record of them for future use), and having transparent information on what is being archived, how it is preserved, and how and when it can be accessed.

This makes e-journals one of the most dynamic and challenging areas of digital preservation, particularly in terms of business models and trust mechanisms for shared or out-sourced preservation services.

Services and solutions

It is important to understand the significant implications for preservation and access of the different requirements (and terminology) that apply for e-journals: in particular the distinction between continuing access and long term preservation, as these differences lead to different types of service for e-journal archiving.

- **Continuing access** (sometimes also called post-cancellation or perpetual access) applies only to subscription journals and securing long-term access for their subscribers;
- **Long-term preservation** applies to both open and subscribed content.

The main preservation and continuing access services and solutions available for e-journals are as follows:

Keepers Registry
The Keepers Registry is a Jisc service to provide easily accessible information about inclusion of e-journals in preservation services and to highlight those e-journals for which no archiving arrangements exist. EDINA, a national data centre based at the University of Edinburgh, has developed the service along with its partner in the project, the ISSN International Centre in Paris
Legal and voluntary deposit in copyright libraries
The role of a national library is to ensure that the published heritage of its country is preserved and made accessible. In many countries legal deposit is an important vehicle for achieving this is. There is a global trend towards extending legal deposit from the print environment to cover e-journals and other electronic publications. Legal deposit legislation (or similar voluntary deposit arrangements) normally involves those subscription e-journals considered part of the national published heritage of that country. To protect the commercial interests of the publisher it also restricts off-site access to preserved electronic material for a substantial period of time. Typically this means a national legal deposit collection does not cover the international range of subscription e-journals licensed by other libraries and their users, and does not meet their requirements for 'perpetual access' rights.
CLOCKSS
CLOCKSS (Controlled LOCKSS) is a not-for-profit collaboration between libraries and publishers. It is a dark archive based on the LOCKSS software (see section below on LOCKSS) in which a limited number of libraries take on an archival role on behalf of a broader community. It provides insurance to libraries that the e-journal and other content they have subscribed to will be preserved for the long term. It is described as a 'private LOCKSS network'.

KB e-Depot
The Koninklijke Bibliotheek (KB) is the national library of the Netherlands and operates the e-Depot. It has taken the policy decision to archive journals that are within its national mandate and additionally a range of e-journals (including open-access titles in the Directory of Open-Access Journals) published beyond its borders. The e-Depot does not currently provide for post-cancellation continuing access by licensees of the content. Generally, end-user access is restricted to on-site perusal at the KB for reasons of private research only and online access is denied. However, full online access is granted to publications by open-access publishers.
LOCKSS
LOCKSS (Lots of Copies Keep Stuff Safe) provides libraries with open-source tools and support so they can take local custody of a wide variety of materials, including subscription and open-access scholarly assets (books, journals, etc.). Readers access LOCKSS preserved content whenever (and for whatever reason) the material cannot be viewed on the publisher's (or intermediary's) servers. The highly distributed nature of this approach aims to ensure that there is sufficient replication to safeguard content despite any potential disasters which might befall individual LOCKSS institutions.
Portico
Portico is designed specifically as a third-party service for scholarly literature published in electronic form and provides three specific preservation services for e-journals, e-books and digitized historical collections respectively. It provides insurance to libraries that the e-journal and other content they have subscribed to will be preserved for the long term. Portico only provides access to the e-journals they have preserved after specified 'trigger events'. In addition, if a publisher has designated Portico as such, it can also serve as a potential mechanism for post-cancellation access.
Consortial hosting
A small number of regional consortia also organize and provide their own hosting services for access and preservation of e-journals. Notable examples are OhioLink, operated by the Ohio Library and Information Network, and the Scholars Portal, operated by the Ontario Council of University Libraries.

Case study 1: the e-journal or its past issues are no longer available from the publisher

This is a highly likely scenario as publishers merge or change their business models, as larger publishers review and adjust their portfolio of titles, or as learned societies move publication contracts for their journals from one publisher to another. Journal titles are also sometimes traded between publishers, which may mean that access to past issues is no longer supported by the previous owner.

The UKSG Transfer Code of Practice initiative has produced a Code of Practice aimed at easing the problems created when journal titles move between publishers. Of relevance are the following paragraphs contained in version 3 of the code (UKSG, 2014):

The transferring publisher will alert the receiving publisher to all existing preservation arrangements for the journal.

The transferring publisher must ensure continued access to its subscribers where it has granted perpetual access rights, even if the transferring publisher will cease to host the online version of the journal after the effective transfer date. Either the transferring or the receiving publisher, or both, could fulfill perpetual access obligations. The Code intentionally does not specify the means for achieving such access, but places on the transferring publisher the responsibility for ensuring that subscribers to whom it has granted perpetual access rights will continue to have access post-transfer

The transferring publisher will use reasonable efforts to communicate journal transfer information where perpetual access rights were granted as part of a licensing agreement/Big Deal, unless archival rights will remain with the transferring publisher.

Subscribers that have been granted perpetual access rights to previously published content with the authority of the journal owner must have those rights honoured. Either the transferring or the receiving publisher, or both, could fulfil perpetual access obligations.

The receiving publisher will continue the existing, or equivalent, preservation arrangements for the journal after the effective transfer date. The receiving publisher will not remove content that was previously deposited in preserving archive(s), even if the receiving publisher will not be continuing to deposit content in the archive(s).

The decision of the publisher Sage to no longer offer its publication Graft provided a real-life example of triggered access from three archiving solutions – Portico, KB e-Depot, and CLOCKSS. In this case all were able to continue to offer access to the issues they held, either as open access (CLOCKSS and KB e-Depot) or else as a service to members (Portico). While it cannot be guaranteed that the archive will include all back issues of the title (as with Graft), participation in an archiving solution which covers at least some issues will significantly reduce the risk of disruption to continuity of service.

Case study 2: library e-Journals, perpetual access, and de-accessioning print

This case-study was first published by Jisc as part of work funded in its digital preservation programme and was incorporated into the Tech Watch Report. It has been adapted for use in the Handbook.

The case study differs from others in illustrating a few of the issues in realizing some of the potential cost savings from e-journals, particularly space savings. Increasingly, academic libraries are investing heavily in e-journals which duplicate their print back-runs. For libraries facing acute pressures on space, one solution to their problem is to dispose of or relegate print back-runs which overlap with their electronic holdings.

The case study focuses on work at Imperial College London Library in providing a database and toolkit for staff making such de-selection decisions (Cooper and Norris, 2007). Imperial established three criteria to determine the sustainability of their e-journals for de-accessioning of print. Their electronic access was classified as sustainable when at least one of the following applied:

- Imperial had perpetual access rights to the content, via the web. Imperial's perpetual access rights were nowhere near as comprehensive as they would have wished; they estimated that less than 50% of their content was covered. In addition, some of their licences specified an unsuitable delivery method for post-termination access. As they were no longer supporting networked CD-ROMs and did not have the resources to mount journal content locally, they considered a journal sustainable only if perpetual access is provided via the web.
- The journal was permanently open access for all years or certain years. Hybrid open-access journals were not included in this category, as the project was not interested in sustainability

at the article level. Finding open-access journals which fulfilled their criteria proved harder than anticipated. The main stumbling block was their need for assurance on the permanency of open access. Although the Bethesda and Berlin Declarations on Open Access include perpetual access in their definitions, Imperial discovered that not all 'open-access journals' met this criterion of permanency.

- The content was in one of Imperial's trusted services such as JSTOR, the ACM digital archive or a Jisc-funded archive. Imperial noted that of their three sustainability criteria, this one, covering services that did not offer perpetual access rights, was the hardest to pin down. The services falling into this category all shared two characteristics: the first was a good track record of stability, i.e., they had demonstrated continuity of titles from one year to another for as long as they had subscribed; the second was a history of and reputation for, affordability and value for money.

Twenty-one months into the project Imperial had identified 700 shelf-metres of sustainable stock for disposal from one site, and planned to rollout the de-selection exercise to other sites. Although it was still early days, they felt their sustainability criteria seemed to be working. The only sustainable content that they had lost was four journals from the same publisher, and they were in the process of challenging that loss. This proved to be an added benefit of the entitlements database they had created for the project; without it they would not have been aware that content over which they had perpetual access rights had been lost.

Conclusions

Continuing access and preservation of e-journals has involved initiatives in organizing multi-institutional collaboration, developing third-party services, and establishing trust in long-term access and preservation between different stakeholders. The issues it has had to address go well beyond technology, and legal, economic and service developments are equally critical to its success. Many challenges remain in e-journal archiving, but there have been significant successes and lessons learnt of interest to the wider digital preservation community as well as to libraries and publishers.

Resources

Preservation, Trust and Continuing Access for e-Journals, DPC Technology Watch Report 13-04 September 2013

http://dx.doi.org/10.7207/twr13-04

This report discusses current developments and issues which libraries, publishers, intermediaries and service providers are facing in the area of digital preservation, trust and continuing access for e-journals. It also includes generic lessons and recommendations on outsourcing and trust learnt in this field of interest to the wider digital preservation community. It is not solely focused on technology, and covers relevant legal, economic and service issues (43 pages).

To bin or not to bin? Deselecting print back-runs available electronically at Imperial College London Library

https://spiral.imperial.ac.uk/handle/10044/1/503

Increasingly, academic libraries are investing heavily in e-journals which duplicate their print back-runs. For libraries facing acute pressures on space, one solution to their problem is to dispose of or relegate print back-runs which overlap with their electronic holdings. This 2007 article by R Cooper and D Norris describes work at Imperial College London Library to provide a tool-kit for staff making such de-selection decisions.

UKSG, 2014 Transfer Code of Practice: Version 3.0 March 2014

http://www.uksg.org/Transfer/Code

The Transfer Code of Practice promotes a set of standards that apply whenever a journal is transferred from one publisher or publishing platform to another. Publishers who publicly sign up to the Code and apply it in practice are considered 'Transfer compliant'. As a voluntary best practices code for industry participants, the Transfer Code of Practice does not supplant contractual terms, intellectual property rights or the competitive marketplace between publishers.

CLOCKSS

http://www.clockss.org

KB e-Depot

http://www.kb.nl/en/organisation/research-expertise/long-term-usability-of-digital-resources/information-for-international-publishers

LOCKSS

http://www.lockss.org

Portico

http://www.portico.org

Ohio Link

http://www.ohiolink.edu

Scholars Portal

http://www.ocul.on.ca/node/135

Keepers Registry

http://thekeepers.org

References

Beagrie, N., 2013. Preservation, Trust and Continuing Access for e-Journals *DPC Technology Watch Report* 13-04 September 2013. Available: http://dx.doi.org/10.7207/twr13-04

Cooper, R. and Norris, D., 2007. *To bin or not to bin? Deselecting print back-runs available electronically at Imperial College London Library*, Serials 20 (3), 208–214. Available: https://spiral.imperial.ac.uk/handle/10044/1/503

UKSG, 2014. *Transfer Code of Practice*: Version 3.0 March 2014. Available: http://www.uksg.org/Transfer/Code

Moving pictures and sound

Illustration by Jørgen Stamp digitalbevaring.dk CC BY 2.5 Denmark

Overview

This case study provides a brief novice to intermediate level overview summarised from the DPC Technology Watch Report on Preserving Moving Picture and Sound. Five "mini case studies" of UK collections that have run preservation and access projects for sound and moving image content are included. The report itself provides a "deep dive" discussing a wider range of issues and practice in greater depth with extensive further reading and advice (Wright, 2012). It is recommended to readers who need a more advanced level briefing on the topic and practice.

Introduction

The audiovisual domain is unique in that digitization is routinely critical to preservation. Audiovisual digitization for preservation is so pervasive that the two words have come to be used interchangeably. Audio and video need digitization for the very survival of their content, owing to the obsolescence of playback equipment and decay and damage of physical items, whether analogue or digital. The basic technology issue for collections of moving images and sound is the necessity to digitize all content currently sitting on shelves. Film on shelves can be conserved (unless it is already deteriorating), but still needs digitization to provide access.

A vital issue in preservation is access: motivation and funding for digitization purely for preservation purposes is difficult, if not impossible. There is great public, institutional and educational interest in the audiovisual record of the twentieth century. Creating access to that record is the key to obtaining the support needed for the digitization and preservation of the content.

The landscape for 'moving pictures and sound' is complicated: physically, there are large differences between audio, video and film recordings. The formats and record/playback equipment are completely separate; the digitization procedures are different; the digital files have different wrapper formats and metadata (with some overlaps); and the storage requirements differ, with video taking roughly 100 times as much storage per second of material as does audio, and high resolution digital film taking roughly 10 times more storage than video.

In addition culturally and economically, there are significant preservation and curation differences between collections from:

- **commercial media industries** – music, cinema and commercial broadcasting where preservation needs a commercial justification, a business case;
- **public bodies** – public service broadcasting, academic collections and heritage institutions such as national museums, libraries and film institutes where preservation needs a cultural heritage justification, though increasingly this sector also needs a business case;
- **technical areas** such as medicine, geology and surveillance, where recordings of images or of seismic events are raw data, kept as medical records or for reprocessing; and
- **other** – a wide range of independent collections, ranging from individual efforts to material gathered by non-profit specialist institutions (for example, steam engine clubs or ethnological research) that do not fall into any of the above categories, though their material may eventually end up being donated to a public collection.

Within the landscape is a range of technologies including engineering, computing, Internet technology, archiving, media management, museum collections management, curation, preservation, access, knowledge management and resource discovery.

Technical challenges

Audiovisual recordings are surrogate reality. The technology allows the listener and viewer to get a sensation of what a situation sounded and looked like, but the technology actually only captures the sequence of light patterns or sound pressures acting on the recording instrument (camera, microphone). These patterns (for film) and signals (for video and audio) are more like data than like artefacts. The preservation requirement is not to keep the original recording media, but to keep the data, the information, recovered from that media.

A key technology issue is moving digital content from carriers (such as CD and DVD, digital videotape, DAT and minidisc) into files. This digital to digital 'ripping' of content is an area of digital preservation unique to the audiovisual world, and has unsolved problems of control of errors in the ripping and transfer process.

The final technology area is digital preservation of the content within the files that result from digitization or ripping, and the files that are born digital. While much of this preservation has problems and solutions in common with other content, there is a specific problem of preserving the quality of the digitized signal that is again unique to audiovisual content. Managing quality through cycles of lossy encoding, decoding and reformatting is one major digital preservation challenge for audiovisual files. The other issue is managing embedded metadata.

For three decades for audio, and for at least two decades for video, archives have been digitizing their analogue content for preservation and access. The problem areas are:

- successful playback of the originals, in order to get an optimal signal to digitize;

- standards: what compression level, encoding method and file format to use; and
- efficiency: digitizing the existing analogue materials fast enough and economically enough to cope with the size and urgency of the problem.

Stages in sound and moving image digital preservation

For sound and moving image preservation, the following stages in the overall process need to be kept clear:

- **signal**: the audio from a microphone, the video signal coming out of a video camera. These signals have physical properties (bandwidth; dynamic range) that can be defined and measured. The quality of a recording and the success or failure of any process of copying, digitization or preservation can be reduced (in large part) to how well that process maintains these two physical properties of the original signal;
- **recording of a signal onto a carrier** (also called support, physical medium or recording format). For a century, the methods of capturing a signal were tied to the carrier of the signal: a wax cylinder, film reel or videotape. Digital technology produces recordings that are independent of carriers. Carrier independence is liberation: discs, tapes and films deteriorate or get damaged. Born digital recordings are liberated from these carrier-based problems, leading to a desire to liberate analogue recordings by digitization;
- **digitization**: analogue recordings can be played back and recorded onto a new carrier, or digitized and so released from carrier dependence. Digitization has to ensure that the digital version has the same bandwidth and dynamic range as the original, to capture the original quality; and
- **digital preservation of the digital representation of a signal**, meaning preserving the numbers, but also preserving the technology needed to decode (render) the numbers. Audiovisual content has a particular problem. The coding of the signal can be a compromise, not actually capturing the full signal, but instead losing some of it (lossy encoding) to get a more compact representation, thus reducing storage and transmission costs. Unfortunately coders/decoders (codecs) go out of use, and are replaced by newer technology. The file format holding the coded signal, the wrapper, is also subject to obsolescence. The failure and obsolescence of storage technology and the obsolescence of encode/decode methods and wrapper formats are major digital preservation problems for audiovisual content.

Access and rights

Sound and moving picture content arising from cinema, broadcasting and the commercial music industry is constrained by rights issues. Music has copyright protection for the composer and for the physical object containing a performance (so-called magnetic copyright). Cinema productions are protected, and music used in a film retains its separate protections. Broadcasting is even more complicated, as all the parties involved in a production may have rights in future exploitation subsequent to the one or two transmissions that were specified in typical contracts. These rights are seen as protection by rights holders, but are also seen as restrictions on access. The situation for a public broadcaster is particularly difficult. The public invariably feel that any production by a public broadcaster has already been paid for by them, is already publicly owned and should be available for public access. Unfortunately that understandable feeling is not the same as the legal definition governing when a work enters the public domain (usually determined by expiry dates on copyright and other rights).

Case study 1: The Open University (OU) Access to video assets project

This is an access and re-use project. The focus is to digitize (where necessary) audiovisual assets previously created by the OU, and place them in an asset management system so that current OU teaching and other activity can find and use these assets. Preservation is a by-product of the project rather than an end in itself. This project provides an important example of combining preservation of content with use of content, something of value to the institution in order to obtain a budget and deliver a benefit. The project was presented at the DPC Briefing Day 'Preserving Digital Sound and Vision'. The project digitized 1,200 videotapes and films, and placed the results in a Fedora digital repository. Also, 145,000 pages of documentation were digitized, providing the overall educational framework around the 1,200 items, giving them context and enhancing their ability to be re-used. The user interface provides granularity and time-based navigation. Overall this project is an outstanding example of best practice.

Case study 2: British Library Archival sound recordings project

This is a JISC-supported preservation and educational access project that ran (in its initial phase) from 2004 to 2006. A second phase added further material. Nearly 50,000 recordings of speech, music and sounds of 'human and natural environments' were digitized and placed online. The online catalogue is open to all and licensed UK further or higher education institutions can also listen to the audio. Anyone can listen to 2,000 of the items (or any of them by attending the British Library reading room in London). The differences in access between educational institutions and the general public reflects the overall issue of rights as the one remaining constraint on open access to audiovisual materials in public institutions.

Case Study 3: Imperial War Museum PSRE project

The Imperial War Museum has one of the UK's major film collections. It has been collecting film since its founding in 1919, beginning with footage from the Great War that led to the institution's founding. The Public Sector Research Exploitation (PSRE) fund made an award of nearly £1 million for cataloguing, digitization and online access (to the catalogue and the footage). The project ran from 2006 to 2009 and is of particular interest in that it is specifically aimed at commercial exploitation of a collection, and at sustainable business models around digitization and web access. The result is a website (http://film.iwmcollections.org.uk/) where anyone can view content in low quality; pull documents, stills and key frames into a lightbox; and fill a shopping basket to then purchase content.

Case Study 4: British University Film and Video Council Newsfilm Online project

This is another project with JISC sponsorship. For four decades to 1960 newsreels shown in cinemas were the main way for the general public to see moving images of current events. The initial project ran from 2004 to 2008. The results are available through a website which, as for the BL Archival Sound Recordings project, has full functionality for registered universities and colleges. The general public can see the full catalogue and can see a single key frame for each item. Since the original phase of the project, the content has been augmented by ITN/Reuters news covering the events from decades after the decline of newsreels. Newsreel items are short: the initial project provided 3,000 hours of content, but that represented 60,000 items. In addition, as with the Open University project, documentation was also placed online for context and to support search and retrieval: 450,000 pages of bulletin scripts.

Case Study 5: BFI and Regional Film Archives Screen Heritage UK (SHUK) project

SHUK is a large (£22.8 million) and complex project (involving 12 regional film archives in addition to the BFI). The project was complicated by changes in the structure and funding of the BFI, as well as a change of government and a raft of other issues. Nevertheless the project has produced major achievements:

- conservation, not digitization: construction of a £6-million vault for film conservation;
- digitization: film scanning and digital storage equipment for the regional film archives;
- access: online catalogues of regional film archive content, available to the general public.

SHUK launched on 5 September 2011 with a BBC BFI joint production, *The Reel History of Britain* (SHUK, 2011).

Conclusions

The basic technology issue for collections of moving images and sound is the necessity for digitization of all content that is currently sitting on shelves. Audio and video need digitization for their very survival, owing to obsolescence and decay of physical items, whether analogue or digital. Film on shelves can be conserved (unless it is already deteriorating) but needs digitization for access.

Playback for preservation-quality digitization implies the need for optimal recovery of the original quality, which requires professional equipment and experience. The major technical obstacle is that, for many physical formats, the needed equipment is largely obsolete, meaning that parts and repairs and skilled operators are in increasingly short supply. The urgent recommendation is, do not wait! Audiovisual holdings need to be documented and made part of a preservation plan.

The situation for sound heritage is clear. The digitization standards, encoding, wrapper and metadata are all agreed and well documented in IASA TC-04 (IASA, 2009). Uncompressed audio in the Broadcast Wave Format (BWF) wrapper is widely used and well supported. There is no reason for the basic encoding to ever be changed, though the BWF wrapper may eventually become obsolete. The only significant problem is the failure of some standard audio applications to handle embedded BWF metadata correctly (ARSC, 2011). All archives need to be aware of the risk of loss of embedded metadata. The situation for video is complex, but there is a PrestoSpace roadmap for guiding choices on the digitization of various legacy formats. There is advice from the PrestoCentre and from JISC Digital Media on the digital preservation of the resultant files. A big challenge is a registry of applications that work properly on embedded video metadata, where the diversity is huge. There is no single agreed wrapper, metadata standard or even encoding standard, and the change from standard definition to high definition brings a new set of applications, wrappers and encodings.

There is emerging technology that can improve audio (capture of the bias tone and consequent removal of temporal variation) and video transfers (direct digitization of the RF signal from the read head), which could be useful in those cases where current technology fails. So the recommendation is not to wait until such technology is further advanced and more widely available. If there are playback problems that cannot be resolved, the original audio or video format should be kept so that such advanced technology can be applied in the future.

Quality checking of the results of digitization remains an issue for video. There is a need for effective integration of signal processing technology with human checking in order to produce a really efficient method of quality control within a preservation factory approach. Quality checking is equally relevant to digital preservation – any changes or migrations due to digital obsolescence need to be checked for preservation of signal quality. Again, a purely manual approach does not scale (to the tens of millions of hours of audiovisual content in European collections), while purely algorithmic substitutes for 'looking and listening' have never been completely successful and remain an area where further research is needed.

Resources

Wright, R., 2012. Preserving Moving Pictures and Sound DPC Technology Watch Report 12-01 March 2012

http://dx.doi.org/10.7207/twr12-02

This report is for anyone with responsibility for collections of sound or moving image content and an interest in preservation of that content. New content is born digital, analogue audio and video need digitization to survive and film requires digitization for access. Consequently, digital preservation will be relevant over time to all these areas. The report concentrates on digitization, encoding, file formats and wrappers, use of compression, obsolescence and what to do about the particular digital preservation problems of sound and moving images (*33 pages*).

SHUK, 2011. Screen Heritage UK Marks new Era for Britain's Film Archives

http://www.bfi.org.uk/sites/bfi.org.uk/files/downloads/bfi-press-release-screen-heritage-uk-marks-a-new-era-for-britains-film-archives-2011-09-01.pdf

BFI Press release. 8 pages

IASA 2009 IASA TC-04, Guidelines on the Production and Preservation of Digital Audio Objects (IASA-TC 04 Second edition 2009) Canberra, IASA.

http://www.iasa-web.org/audio-preservation-tc04

This is the standard guide to digitization of audio, and the sections on metadata and digital storage are of value to all forms of digital media.

Casey, M. and Gordon, B., 2007. Best Practices for Audio Preservation. Bloomington, Indiana University Bloomington.

http://www.dlib.indiana.edu/projects/sounddirections/papersPresent/

Another audio resource (that also includes a range of digitization software tools) comes from the Sound Directions project of Harvard and Indiana Universities: much is also relevant to video digitization. (*160 pages*)

Digital Preservation Coalition Briefing day on Preserving Digital Sound and Vision, April 2011

http://www.dpconline.org/events/details/27-SoundAndVision?xref=26

This DPC briefing day in April 2011 provided a forum to review and debate the latest development in the preservation of digital sound and vision. Seven presentations (including the Open University) are linked from the programme and available to download.

ARSC Technical Committee, 2011. Study of Embedded Metadata Support in Audio Recording Software. Association of Recorded Sound Collections.

http://www.arsc-audio.org/pdf/ARSC_TC_MD_Study.pdf

A study of support for embedded metadata within and across a variety of audio recording software applications. The findings raise serious concerns, particularly for the archiving and preservation communities who rely on embedded metadata for interpretation and management of digital files representing preserved content into the future. (*21 pages*)

AVPreserve

http://www.avpreserve.com/

US based media and information management consulting firm. It website provides a range of resources for AV preservation.

BUFVC NewsFilm online Project

http://www.webarchive.org.uk/wayback/archive/20140614061518/http://www.jisc.ac.uk/whatwedo/programmes/digitisation/bufvc.aspx

British Film Institute

http://www.bfi.org.uk

the British Film Institute can advise on film and also on video – they hold a lot of video, and have a Curator for Television. Its remit is collection and preservation of film and television, and technical advice.

British Library Sound Archive

http://www.bl.uk/nsa

General technical advice on audio preservation is available from the British Library Sound Archive. Its remit is collection and preservation of all forms of audio, and technical advice.

Film Archives UK

http://filmarchives.org.uk

Collection and preservation of general audiovisual content of regional significance in the UK

JISC Digital Media

http://www.jiscdigitalmedia.ac.uk

Advice and training on still images, moving images and sound. This includes their InfoKits for Digital File Formats, Digitisation funding and sustainability, and High Level Digitisation Guide for Audiovisual Resources.

PrestoCentre

http://www.prestocentre.eu

Website provides audiovisual information, resources and advice. Access to most resources on the website requires a member subscription but a number are available to non-numbers.

The Preservation Guide Wiki

http://preservationguide.co.uk/RDWiki/

This audiovisual preservation guide was created for PrestoSpace and the BBC in May 2006. The site is now maintained as part of The Preservation Guide Consultancy. The wiki is in the public domain under a creative commons licence.

Sustaining Consistent Video Presentation

http://www.tate.org.uk/research/publications/sustaining-consistent-video-presentation

This technical paper addresses approaches to identifying and mitigating risks associated with sustaining the consistent presentation of digital video files. Originating from two multi-partnered research projects – Pericles and Presto4U – the paper was commissioned by Tate Research and is intended for those who are actively engaged with the preservation of digital video.

JISC 2009 - Archival Sound Recordings Showreel

https://www.youtube.com/watch?v=KPy9ZqWEHog

Engaging short video on British Library archival sound recordings project published on 22 Jun 2009. (6 mins 11 secs).

Further case studies

Podcasts in the Archives: Archiving Podcasting Content at the University of Michigan

http://files.archivists.org/pubs/CampusCaseStudies/CASE12.pdf

In this Society of American Archivists campus case study Alexis. A. Antracoli, University of Michigan, examines the challenges involved in developing best practices and workflows for archiving and preserving podcasting content. One major issue involved establishing standards of practice for ingest, storage, and access, especially the generation and storage of appropriate descriptive, technical, and

preservation metadata. Another challenge centered around developing the necessary technological infrastructure to support an Open Archives Information System (OAIS)-compliant system. 2010. (14 pages).

References

ARSC Technical Committee, 2011. *Study of Embedded Metadata Support in Audio Recording Software*. Association of Recorded Sound Collections. Available: http://www.arsc-audio.org/pdf/ARSC_TC_MD_Study.pdf

IASA, 2009. *IASA TC-04, Guidelines on the Production and Preservation of Digital Audio Objects*, IASA-TC 04 Second edition 2009, Canberra, IASA. Available: http://www.iasa-web.org/audio-preservation-tc04

SHUK, 2011. *Screen Heritage UK Marks new Era for Britain's Film Archives.* Available: http://www.bfi.org.uk/sites/bfi.org.uk/files/downloads/bfi-press-release-screen-heritage-uk-marks-a-new-era-for-britains-film-archives-2011-09-01.pdf

Wright, R., 2012. Preserving Moving Pictures and Sound *DPC Technology Watch Report* 12-01 March 2012. Available: http://dx.doi.org/10.7207/twr12-02

Web-archiving

Illustration by Jørgen Stamp digitalbevaring.dk CC BY 2.5 Denmark

Overview

This case study provides a brief novice to intermediate level overview summarised from the DPC Technology Watch Report on Web-Archiving. Three "mini case studies" are included illustrate the different operational contexts, drivers, and solutions that can be implemented. The report itself provides a "deep dive" discussing a wider range of issues and practice in greater depth with extensive further reading and advice (Pennock, 2013). It is recommended to readers who need a more advanced level briefing on the topic and practice.

Introduction

The World Wide Web is a unique information resource of massive scale, used globally. Much of its content will likely have value not just to the current generation but also to future generations. Yet the lasting legacy of the web is at risk, threatened in part by the very speed at which it has become a success. Content is lost at an alarming rate, risking not just our digital cultural memory but also organizational accountability. In recognition of this, a number of cultural heritage and academic institutions, non-profit organizations and private businesses have explored the issues involved and lead or contribute to development of technical solutions for web archiving.

Services and Solutions

Business needs and available resources are fundamental considerations when selecting appropriate web archiving tools and/or services. Other related issues must also be considered: organizations considering web archiving to meet regulatory requirements must, for example, consider associated issues such as authenticity and integrity, recordkeeping and quality assurance. All organizations will need to consider the issue of selection (i.e. which websites to archive), a seemingly straightforward task which is complicated by the complex inter-relationships shared by most websites that make it difficult to set boundaries. Other issues include managing malware, minimizing duplication of resources, temporal coherence of sites and long-term preservation or sustainability of resources. International collaboration is proving to be a game-changer in developing scalable solutions to support long-term preservation and ensure collections remain reliably accessible for future generations.

The web archiving process is not a one-off action. A suite of applications is typically deployed to support different stages of the process, though they may be integrated into a single end-to-end workflow. Much of the software is available as open source, allowing institutions free access to the source code for use and/or modification at no cost.

Integrated Systems for Web-archiving

A small number of integrated systems are available for those with sufficient technical staff to install, maintain and administer a system in-house. These typically offer integrated web archiving functionality across most of the life cycle, from selection and permissions management to crawling, quality assurance, and access. Three are featured here.

PANDAS
PANDAS (PANDORA Digital Archiving System) was one of the first available integrated web archiving systems. First implemented by the National Library of Australia (NLA) in 2001, PANDAS is a web application written in Java and Perl that provides a user-friendly interface to manage the web archiving workflow. It supports selection, permissions, scheduling, harvests, quality assurance, archiving, and access. PANDAS is not open source software, though it has been used by other institutions (most notably the UK Web Archiving Consortium from 2004 to 2008). It is used by the NLA for selective web archiving, whilst the Internet Archive supports their annual snapshots of the Australian domain.
Web Curator Tool (WCT)
The Web Curator Tool is an open source workflow tool for managing the selective web archiving process, developed collaboratively by the National Library of New Zealand and the British Library with Oakleigh Consulting. It supports selection, permissions, description, harvests, and quality assurance, with a separate access interface. WCT is written in Java within a flexible architecture and is publicly available for download from SourceForge under an Apache public licence. The WCT website is the hub for the developer

community and there are active mailing lists for both users and developers. The highly modular nature of the system minimizes system dependencies.

NetarchiveSuite

NetarchiveSuite is a web archiving application written in Java for managing selective and broad domain web archiving, originally developed in 2004 by the two legal deposit libraries in Denmark (Det Kongelige Bibliotek and Statsbiblioteket). It became open source in 2007 and has received additional development input from the Bibliothèque nationale de France and the Österreichische Nationalbibliothek since 2008. It is freely available under the GNU Lesser General Public License (LGPL). The highly modular nature of the system enables flexible implementation solutions.

Third party and commercial services

Third party commercial web archiving services are increasingly used by organizations that prefer not to establish and maintain their own web archiving technical infrastructure. The reasons behind this can vary widely. Often it is not simply about the scale of the operation or the perceived complexity, but the business need and focus. Many organizations do not wish to invest in any skills or capital that is not core to their business. Others may use such a service to avoid capital investment. Moreover, organizations are increasingly moving their computing and IT operations into the cloud, or using a SAAS (Software as a Service) provider. Web archiving is no exception. From a legal and compliance perspective, third party services are sometimes preferred as they can provide not just the technology but also the skills and support required to meet business needs. This section introduces some of the third party services currently available but is of course a non-exhaustive list, and inclusion here should not be taken as recommendation.

Archive-It

Archive-It is a subscription web archiving service provided by the Internet Archive. Customers use the service to establish specific collections, for example about the London 2012 Olympics, government websites, human rights, and course reading lists. A dedicated user interface is provided for customers to select and manage seeds, set the scope of a crawl and crawl frequency, monitor crawl progress and perform quality assurance, add metadata and create landing pages for their collections. Collections are made public by default via the Archive-It website, with private collections requiring special arrangement. The access interface supports both URL and full text searching. Over 200 partners use the service, mostly from the academic or cultural heritage sectors. The cost of the service depends on the requirements of the collecting institution

Archivethe.Net

Archivethe.Net is a web-based web archiving service provided by the Internet Memory Foundation (IMF). It enables customers to manage the entire workflow via a web interface to three main modules: Administration (managing users), Collection (seed and crawl management), and Report (reports and metrics at different levels). The platform is available in both English and French. Alongside full text searching and collection of multimedia content, it also supports an automated redirection service for live sites. Automated QA tools are being developed though IMF can also provide manual quality assurance services, as well as direct collection management for institutions not wishing to use the online tool. Costs are dependent upon the requirements of the collecting institution. Collections can be made private or remain openly accessible,

in which case they may be branded as required by the collecting institutions and appear in the IMF collection. The hosting fee in such cases is absorbed by IMF.
The University of California's Curation Centre (UC3)
As part of the California Digital Library, provides a fully hosted Web Archiving Service for selective web archive collections. University of California departments and organizations are charged only for storage. Fees are levied for other groups and consortia, comprising an annual service fee plus storage costs. Collections may be made publicly available or kept private. Around 20 partner organizations have made collections available to date. Full text search is provided and presentation of the collections can be branded as required by collecting institutions.
Private companies
Private companies offer web archiving services particularly tailored to business needs. Hanzo Archives, for example, provide a commercial website archiving service to meet commercial business needs around regulatory compliance, e-discovery and records management. Hanzo Archives emphasize their ability to collect rich media sites and content that may be difficult for a standard crawler to pick up, including dynamic content from Sharepoint, and wikis from private internets, alongside public and private social media channels. (More details about the possibilities afforded by the Hanzo Archives service can be found in the Coca-Cola case study) Similarly, Reed Archives provide a commercial web archiving service for organizational regulatory compliance, litigation protection, eDiscovery and records management. This includes an 'archive-on-demand' toolset for use when browsing the web. In each case, the cost of the service is tailored to the precise requirements of the customer. Other companies and services are also available and readers are encouraged to search online for further options should such a service be of interest.

Case study 1: The UK Web Archive

The UK Web Archive (UKWA) was established in 2004 by the UK Web Archiving Consortium. It was originally a six-way partnership, led by the British Library in conjunction with the Wellcome Library, Jisc, the National Library of Wales, the National Library of Scotland and The National Archives (UK).

UKWA partners select and nominate websites using the features of the web archiving system hosted on the UK Web Archive infrastructure maintained by the British Library. The British Library works closely with a number of other institutions and individuals to select and nominate websites of interest. Selectively archived websites are revisited at regular intervals so that changes over time are captured.

The technical infrastructure underpinning the UK Web Archive is managed by the British Library. The Archive was originally established with the PANDAS software provided by the National Library of Australia, hosted by an external agency, but in 2008 the archive was moved in-house and migrated into the Web Curator Tool (WCT) system.

A customized version of the Wayback interface developed by the Internet Archive is used as the WCT front end and provides searchable access to all publicly available archived websites. Full text searching is enabled in addition to standard title and URL searches and a subject classification schema. The web archiving team at the library have recently released a number of visualization tools to aid researchers in understanding and finding content in the collection.

Special collections have been established on a broad range of topics. Many are subject based, for example the mental health and the Free Church collections. Others document the online response to a notable event in recent history, such as the UK General Elections, Queen Elizabeth II's Diamond Jubilee and the London 2012 Olympics.

Many more single sites, not associated with a given special collection, have been archived on the recommendation of subject specialists or members of the public. These are often no longer available on the live web, for example the website of UK Member of Parliament Robin Cook or Antony Gormley's One & Other public art project , acquired from Sky Arts.

Case study 2: The Internet Memory Foundation

The Internet Memory Foundation (IMF) was established in 2004 as a non-profit organization to support web archiving initiatives and develop support for web preservation in Europe. Originally known as the European Archive Foundation, it changed its name in 2010. IMF provides customers with an outsourced fully fledged web archiving solution to manage the web archiving workflow without them having to deal with operational workflow issues.

IMF collaborates closely with Internet Memory Research (IMR) to operate a part of its technical workflows for web archiving. IMR was established in 2011 as a spin off from the IMF. Both IMF and IMR are involved in research projects that support the growth and use of web archives.

IMR provides a customizable web archiving service, Archivethe.Net (AtN). AtN is a shared web-archiving platform with a web-based interface that helps institutions to easily and quickly start collecting websites including dynamic content and rich media. It can be tailored to the needs of clients, and institutions retain full control of their collection policy (ability to select sites, specify depth, gathering frequency, etc.). Quality control services can be provided on request. Most is done manually in order to meet high levels of institutional quality requirements, and IM has a dedicated QA team composed of QA assessors. IM has developed a methodology for visual comparison based on tools used for crawling and accessing data, though they are also working on improving tools and methods to deliver a higher initial crawl quality.

Partner institutions, with openly accessible collections for which the IM provides a web archiving service, include the UK National Archives and the UK Parliament.

Access to publicly available collections is provided via the IM website. IM provides a full text search facility for most of its online collections, in addition to URL-based search. Full text search results can be integrated on a third party website and collections can be branded by owners as necessary.

Following the architecture of the Web Continuity Service by The National Archives (The National Archives, 2010), IM implemented an 'automatic redirection service' to integrate web archives with the live web user experience. When navigating on the web, users are automatically redirected to the web archive if the resource requested is no longer available online. Within the web archive, the user is pointed to the most recent crawled instance of the requested resource. Once the resource is accessed, any link on the page will send the user back to the live version of the site. This service is considered to increase the life of a link, to improve users' experience, online visibility and ranking, and to reduce bounce rates.

Web archiving collections are available for public browsing from the IM website, a combination of both domain and selective collections from its own and from partner institutions.

Case study 3: The Coca-Cola web archive

The Coca-Cola Web Archive was established to capture and preserve corporate Coca-Cola websites and social media. It is part of the Coca-Cola Archive, which contains millions of both physical and digital artefacts, from papers and photographs to adverts, bottles, and promotional goods. Coca-Cola's online presence is vast, including not only several national Coca-Cola websites but also for example, the Coca-Cola Facebook page and Twitter stream, and other Coca-Cola owned brands (500 in all).The first Coca-Cola website was published in 1995.

Since 2009, Coca-Cola has collaborated with Hanzo Archives and now utilizes their commercial web archiving service. Alongside the heritage benefits of the web archive, the service also provides litigation support where part or all of the website may be called upon as evidence in court and regulatory compliance for records management applications.

The Coca-Cola web archive is a special themed web archive that contains all corporate Coca-Cola sites and other specially selected sites associated with Coca-Cola. It is intended to be as comprehensive as possible, with integrity/functionality of captured sites of prime importance. This includes social media and video, whether live-streamed or embedded (including Flash). Artefacts are preserved in their original form wherever possible, a fundamental principle for all objects in the Coca-Cola Archive.

Hanzo Archives' crawls take place quarterly and are supplemented by occasional event-based collection crawls, such as the 125th anniversary of Coca-Cola, celebrated in 2011. Hanzo's web archiving solution is a custom-built application. Web content is collected in its native format by the Hanzo Archives web crawler, which is deployed to the scale necessary for the task in hand.

Quality assurance is carried out with a two-hop systematic sample check of crawl contents that forces use of the upper-level navigation options and focuses on the technical shape of the site.

The Archive is currently accessible only to Coca-Cola employees, on a limited number of machines. Remote access is provided by Hanzo using their own access interface. Proxy-based access ensures that all content is served directly from the archive and that no 'live-site leakage' is encountered. The archive may be made publicly accessible in the future inside The World of Coca-Cola, in Altanta, Georgia, USA.

The Coca-Cola web archive collection contains over six million webpages and over 2TB of data. Prior to their collaboration with Hanzo, early attempts at archiving resulted in incomplete captures so early sites are not as complete as the company would like. The collection also contains information about many national and international events for which Cola-Cola was a sponsor, including the London 2012 Olympics and Queen Elizabeth II's Diamond Jubilee.

Conclusions

Web archiving technology has significantly matured over the past decade, as has our understanding of the issues involved. Consequently we have a broad set of tools and services which enable us to archive and preserve aspects of our online cultural memory and comply with regulatory requirements for capturing and preserving online records. The work is ongoing, for as long as the Internet continues to evolve, web archiving technology must evolve to keep pace.

Alongside technical developments, the knowledge and experience gained through practical deployment and use of web archiving tools has led to a much better understanding of best practices in web archiving, operational strategies for embedding web archiving in an organizational context, business needs and benefits, use cases, and resourcing options. Organizations wishing to embark on a web archiving initiative must be very clear about their business needs before doing so. Business needs

should be the fundamental driver behind any web archiving initiative and will significantly influence the detail of a resulting web archiving strategy and selection policy. The fact that commercial services and technologies have emerged is a sign of the maturity of web archiving as a business need, as well as a discipline.

Resources

Pennock, M., 2013. Web-Archiving, DPC Technology Watch Report 13-01 March 2013

http://dx.doi.org/10.7207/twr13-01

This report is intended for those with an interest in, or responsibility for, setting up a web archive. It introduces and discusses the key issues faced by organizations engaged in web archiving initiatives, whether they are contracting out to a third party service provider or managing the process in-house and provides a detailed overview of the main software applications and tools currently available.

ISO, 2012, ISO 28500:2009 Information and Documentation – the WARC file format

http://www.iso.org/iso/catalogue_detail.htm?csnumber=44717

The WARC (Web ARChive) format is a container format for archived websites, also known as ISO 28500:2009. It is a revision of the Internet Archive's ARC File Format used to store web crawls harvested from the World Wide Web.

ISO, 2013 ISO/TR 14873:2013 Information and Documentation – Statistics and quality issues for web archiving

http://www.iso.org/iso/catalogue_detail.htm?csnumber=55211

This technical report defines statistics, terms and quality criteria for Web archiving. It considers the needs and practices across a wide range of organisations such as libraries, archives, museums, research centres and heritage foundations.

Meyer E 2010 (a), Researcher Engagement with Web Archives: State of the Art Report, JISC

http://ie-repository.jisc.ac.uk/544/

This report summarizes the state of the art of web archiving in relationship to researchers and research needs focussing primarily on individual researchers and institutions.

Perma: Scoping and Addressing the Problem of Link and Reference Rot in Legal Citations

Zittrain, Jonathan and Albert, Kendra and Lessig, Lawrence, Perma: Scoping and Addressing the Problem of Link and Reference Rot in Legal Citations (October 1, 2013). Harvard Public Law Working Paper No. 13-42. Available at SSRN: http://ssrn.com/abstract=2329161 or http://dx.doi.org/10.2139/ssrn.2329161

http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2329161 or http://dx.doi.org/10.2139/ssrn.2329161

This article from the Perma project team documents a serious problem of reference rot: more than 70% of the URLs within the Harvard Law Review and other journals, and 50% of the URLs found within United States Supreme Court opinions, do not link to the originally cited information. It proposes a solution for authors and editors of new scholarship that involves libraries undertaking the distributed, long-term preservation of link contents.

Scholarly Context Not Found: One in Five Articles Suffers from Reference Rot

http://dx.doi.org/10.1371/journal.pone.0115253

This large-scale study looked into approximately 600K links extracted from over 3M scholarly papers published between 1997 and 2012. Those were links to so-called web-at-large resources, i.e. not links to other scholarly papers. It found one out of five STM articles suffering from reference rot, meaning it is impossible to revisit the web context that surrounds them some time after their publication. When only considering STM articles that contain references to web resources, this fraction increases to seven out of ten.

The National Archives, 2010.Government Web Archive: Redirection Technical Guidance for Government Departments, version 4.2, The National Archives (UK)

http://www.nationalarchives.gov.uk/documents/information-management/redirection-technical-guidance-for-departments-v4.2-web-version.pdf

This guidance describes an innovative service that provides URL rewriting and redirection functionality for UK Government web pages by setting up redirection to the UK Government web archive where a requested URL does no longer exists on a departmental web site.

MEMENTO and the Time Travel Service

http://www.mementoweb.org/

Memento is a tool which allows users to see a version of a web resource as it existed at a certain point in the past. It is now used in several web archives. The Time Travel service based on Memento checks a range of servers including many web archives and tries to find a web page as it existed around the time of your choice.

Archive-It

http://www.archive-it.org/

Archivethe.Net

http://www.archivethe.net/en/

Hanzo Archives

http://www.hanzoarchives.com/

Internet Memory Foundation & Internet Memory Research

http://www.internetmemory.org/en/

Wayback

http://www.sourceforge.net/projects/archive-access/files/wayback/

Netarchive Suite

https://sbforge.org/display/NAS/NetarchiveSuite

PANDAS

http://pandora.nla.gov.au/pandas.html

Reed Archives

http://www.reedarchives.com/

UC3 Web Archiving Service

http://www.cdlib.org/services/uc3/was.html

Web Curator Tool

http://webcurator.sourceforge.net/

International Internet Preservation Consortium

http://www.netpreserve.org

The IIPC is a membership organization dedicated to improving the tools, standards and best practices of web archiving while promoting international collaboration and the broad access and use of web archives for research and cultural heritage. There are many valuable resources on the website including excellent short videos such as the example below.

Why Archive the Web?

https://www.youtube.com/watch?v=pU32rjTaMFE

A short video published on 18 Oct 2012 introducing the challenges of web-archiving and the IIPC. (2 mins 53 secs).

What is a Web Archive?

https://youtu.be/ubDHY-ynWi0

This short video explains 'Web Archiving' and why is it important that the UK Legal Deposit libraries support it. It was produced as part of the Arts and Humanities Research Council funded 'Big UK Domain Data for the Arts and Humanities' project.(2 mins 31 secs)

What do the UK Web Archive collect?

https://youtu.be/1QLMPIRwJEo

This video for users explains what they can expect to find and where they might go to access the three collections that the UK Web Archive hold. It was produced as part of the Arts and Humanities Research Council funded 'Big UK Domain Data for the Arts and Humanities' project. (2 mins 55 secs)

Further case studies

NDSA Website content case studies

The US National Digital Stewardship Alliance (NDSA) examines the value, opportunities and obstacles for selective preservation of the following specific web content types:

Science, Medicine, Mathematics, and Technology forums

http://www.digitalpreservation.gov/ndsa/working_groups/documents/ScienceForums_CaseStudy_public_v2.pdf

December 2013 (3 pages).

Science, Medicine, Mathematics, and Technology blogs

http://www.digitalpreservation.gov/ndsa/working_groups/documents/ScienceBlogs_CaseStudy_public_v2.pdf

December 2013 (3 pages).

Born-Digital Community and Hyperlocal News

http://www.digitalpreservation.gov/ndsa/working_groups/documents/NDSA_CaseStudy_CommunityNews.pdf

February 2013 (3 pages).

Citizen Journalism

http://www.digitalpreservation.gov/ndsa/working_groups/documents/NDSA_CaseStudy_CitizenJournalism.pdf

February 2013 (3 pages).

On the Development of the University of Michigan Web Archives: Archival Principles and Strategies

http://files.archivists.org/pubs/CampusCaseStudies/Case13Final.pdf

Michael Shallcross, Bentley Historical Library, University of Michigan details the strategies and procedures the University Archives and Records Program (UARP) followed to develop its collection of archived websites, and how it initiated a large-scale website preservation project as part of a broader effort to proactively capture and maintain select electronic records of the University. 2011 (29 pages).

References

Pennock, M., 2013. Web-Archiving, *DPC Technology Watch Report* 13-01 March 2013. Available: http://dx.doi.org/10.7207/twr13-01

The National Archives, 2010. *Government Web Archive: Redirection Technical Guidance for Government Departments*, version 4.2, The National Archives (UK). Available: http://www.nationalarchives.gov.uk/documents/information-management/redirection-technical-guidance-for-departments-v4.2-web-version.pdf