Fake news and dark ads: Why digital preservation has never mattered more

Richard Ovenden
Bodley’s Librarian
BREAKING NEWS
TRUMP: "I THINK IT WAS RUSSIA" BEHIND ELECTION HACKS
Jake Tapper | CNN Chief Washington Correspondent
Link Rot

- Zittrain, Lessig, Albert 2014 study found that 50% of links in US Supreme Court website were broken.
- They also looked at a selection of legal journals 1999–2011 and found that 70% of the links no longer functioned as intended.
- A 2013 study in BMC Bioinformatics analysed 15k links in abstracts in Web of Science and found median lifespan of web pages was 9 years and 62%.
- In 2015 Weblock analysed 180k references in full-text corpora of three open-access publishers and found 24% of the links were not available.
PERMA: SCOPING AND ADDRESSING THE PROBLEM OF LINK AND REFERENCE ROT IN LEGAL CITATIONS

Jonathan Zittrain, Kendra Albert, and Lawrence Lessig

INTRODUCTION

Works of scholarship have long cited primary sources or academic works to provide sources for facts, to incorporate previous scholarship, and to bolster arguments. The ideal citation connects an interested reader to what the author references, making it easy to track down, verify, and learn more from the indicated sources.

In principle, as cited sources move to the Web, this linking should become easier. Rather than requiring a reader to travel to a library to follow the sources cited by an author, the reader should be able to retrieve the cited material immediately with a single click.

But again, only in principle. The link, a URL, points to a resource hosted by a third party. That resource will only survive so as long as the third party preserves it. And as websites evolve, not all third parties will have a sufficient interest in preserving the links that provide backwards compatibility to those who relied upon those links. The author of the cited source may decide the argument in the source was mistaken and take it down. The website owner may decide to abandon one mode of organizing material for another. Or the organization providing the source material may change its views and “update” the


The Content Explosion

“Every two days we create as much content as from the dawn of man up to the year 2003.”

- Eric Schmidt, Google, in 2010
DP Until now

Smelling the coffee
Awareness of loss
Awareness that we needed to do something
Reports, studies projects
Standards: OIAS, TIFF, PDF
Projects: CEDARS, RLG,
DP Until now

Drinking the coffee

Doing stuff: LOCKSS, CLOCKSS, NDIIP ...

Services: PRONOM ...

Tools: JHOVE, DROID

Organising ourselves: DPC, Open Preservation Foundation, DPN

Internet Archive, UKWebarchive, Memento

Institutional repositories

Shared ventures: Hathi Trust
Email functionality temporarily disabled

Please note that for security reasons the ability to email records from SOLO has been temporarily disabled. This functionality will be restored as soon as possible and we apologise for any inconvenience caused.

SOLO Live Help

Live Help is offline. Please use our Ask A Librarian service.

You can also text us on: 02030953910

Monday-Friday, 9am-5pm

Latest news

War, empire and business: an exhibition on Jane Austen's unexplored life

Last chance to view the Bodleian's explosive Volcanoes exhibition

Bodleian display showcases scientific research into Bauer's botanical masterpieces

Novelist William Boyd awarded Bodleian's highest honour at Oxford Literary Festival
Schildlied
Andante con moto

Fluss

Aus dem Lus' dem ruh'gem Schwim' und dem Monde's fliess

Sempre piano


Cresc.
The Cedars Project

Digital Curation:
digital archives, libraries, and e-science.

19th October 2001
Forbidden

You don't have permission to access /metadata/cedars/ on this server.

Apache/2.4.18 (Unix) Server at www.ukoln.ac.uk Port 80
404 Page not found

The page you are looking for could not be found. Try searching below.

Search by keyword

cedars index

Search the site

GO
The LOCKSS Program, based at Stanford University Libraries, provides libraries and publishers with award-winning, low-cost, open source digital preservation tools to preserve and provide access to persistent and authoritative digital content.

LIBRARIES KEEP WHAT THEY BUY. Librarians need to know that the digital content they acquire today will not disappear when they cancel subscriptions, and that their electronic collections can be preserved and accessed by readers far into the future.

PUBLISHERS PRESERVE WHAT THEY PUBLISH. Publishers need to know that the integrity of their web-based content will remain unchanged and available in perpetuity—even if their own website is no longer available.
The CLOCKSS Archive
A Trusted Community-Governed Archive

CLOCKSS Announces the Support of Additional Publishers and Libraries

The CLOCKSS Archive announces the trigger of two journals for open access

CLOCKSS Endorses the Statement “Working Together to Ensure the Future of the Digital Scholarly Record

Public Knowledge Project Announces Private LOCKSS Network

CLOCKSS Makes Guidelines Available:

Web Harvest Guidelines
  - CLOCKSS Web Harvest Guidelines

File Transfer Guidelines
  - CLOCKSS File Transfer Guidelines

Our Mission

...let us save what remains: not by vaults and locks which fence them from the public eye and use in consigning them to the waste of time, but by
Services

E-JOURNAL PRESERVATION SERVICE
Since 2005, we have worked with publishers and libraries to preserve a rapidly increasing number of e-journals through this service. Learn more »

E-BOOK PRESERVATION SERVICE

D-COLLECTIONS PRESERVATION SERVICE

Our Preservation Approach
Personal Digital Archives

• Personal behaviour increasingly moving to digital.
• Lives of people prominent in their fields are now documented in hybrid personal archives
• Increase in complexity, as digital materials are spread around numerous devices, and locations.
• Archivists cannot rely on ambulance chasing!
• BUT Issues around FOI driving behaviour of some creators back to analogue.
• Social media also playing a role here …
CHAPTER ONE
EARLY INFLUENCES

My earliest memory is of sitting on the back steps of a square little house, Beechurst, in Love Lane, Pontefract. I must have been three years old at the time and I was looking out at a stretch of unkempt grass surrounded by a brick wall which we called a garden. On re-visiting the house many years later I found it was a very modest home, but to me it represented solidity and security. Beyond our wall was a cinder track which separated us from a terraced row of houses occupied mainly by miners from the surrounding pits. I remember because we used to play with the miners' children, learning some colourful language from them. My mother used to recount how I came in one day from a rough and tumble on the cinder track and, pointing to my elder brother, Jimmie, lisped: "Look at icka Dimmy. Isn't he a mucky little buggar."

At this stage my mother was the dominant influence in my life. She was a William Morris Socialist - wilderness must be turned...
Welcome to paradigm

The Personal Archives Accessible in Digital Media (paradigm) project saw the major research libraries of the Universities of Oxford and Manchester come together to explore the issues involved in preserving digital private papers through gaining practical experience in accessioning and ingesting digital private papers into digital repositories, and processing these in line with archival and digital preservation requirements.

The project was funded between 2005 and 2007, under the JISC's programme "Supporting Institutional Digital Preservation and Asset Management".

Project outcomes

The project outcomes included:

- A template for ensuring long-term access for institutional holdings of digital personal papers
- Best-practice guidelines in the form of a workbook on issues relating to the archiving of personal papers in digital form, made available in sections as they are completed
- Strengthened local institutional capacity for digital preservation
- Developed templates for institutional policies for collection development, retention, and disposal
- Practical test of digital repository softwares DSpace and Fedora and related tools
- Research resources for 20th century political history in the form of new archival collections

Visit the About section to find out more.

[Top]
Web Archives

• Becoming mainstream activity
• Come a long way since CORC!
• Many institutions now include it as mainstream library activity
• Greatly facilitated by tools such as Archive-IT
Why archive the web?

“Web Archiving operates at the frontier of capturing and preserving our contemporary cultural and historical record. From the Information Highway to social networking sites, the Internet represents not only our cultural record but the inscription of an evolving technology”

Sara Day Thomson, Digital Preservation Coalition
Science, medicine and technology

Collected by: Bodleian Libraries

Archived since: Mar, 2011

Description: Web sites relating to science, medicine and technology.

Subject: Computers & Technology, Science & Health, Blogs & Social Media

Narrow Your Results

Sites for this collection are listed below. Narrow your results at left, or enter a search query below to find a site, specific URL or to search the text of archived webpages.

Enter search terms here

Search

Page 1 of 1 (25 Total Results)

Title: CN Bio Innovations

URL: http://cn-bio.com/

Captured on May 5, 2017
Open UK Web Archive

- Selective archiving since 2004
- 15,000 sites
- Quality Assurance and permissions clearing
- All six UK Legal deposit libraries
- Plus many collaborators: Women’s Library, Live Art Development Agency, NHS, HLF

Accessible via: http://webarchive.org.uk
UK Legal Deposit Web Archive

We collect

- .uk and other UK geographic top-level domains, i.e. .london, .scot, .cymru, .wales
- Websites hosted in the UK
- Website owner has UK address

Out of scope

- Audio-visual only content
- Personal data on private intranets/ shared with private groups only
## Numbers and size

<table>
<thead>
<tr>
<th>Open UK Web Archive</th>
<th>Legal Deposit Web Archive</th>
</tr>
</thead>
<tbody>
<tr>
<td>15,000 websites (80,000 instances)</td>
<td>10 million websites</td>
</tr>
<tr>
<td></td>
<td>6 billion objects</td>
</tr>
<tr>
<td>28 TB data</td>
<td>150 TB data (growing at 60 TB per year)</td>
</tr>
<tr>
<td>1.2 million page views 2015/16</td>
<td>32,000 page views 2015/16</td>
</tr>
</tbody>
</table>
Why archive the web?
Bodleian Library
University of Oxford

Announcements:

- Closure of Duke Humfrey's Library 1998/9 Amended Notice

Information about the Bodleian Library
General information about the Bodleian Library and Guides for readers.

Local Applications
Some interesting WWW applications being developed by the Bodleian Library

Telnet to OLIS
Telnet directly to OLIS, the catalogue of the Bodleian and other Oxford Libraries.

OLIG
Oxford Libraries Internet Gateway (OLIG) - supersedes BARD. Instead of attempting to provide its own comprehensive guidance to the Internet, the new service will concentrate on directing users to major services provided on the Internet by providers funded to supply such services

World Wide Internet Resources
A miscellany of links we hope will be of interest to Oxford users exploring the Internet.

Bodleian Database Network
Details about locally mounted CD-ROM and magnetic databases and how to access them. These are only available within Oxford to members of the University.

The Bodleian Shopping Arcade
A selection of attractive items available by mail order from the Bodleian Shop.

Official Library
The Bodleian Library welcomes enquiries, comments and suggestions. This document tells you how to contact the various parts of the Library.
...and a wonderful time was had by all!

We're still awaiting photographic evidence from some of our candidates, but in the meantime, here are a few pictures from campaigns, and from election night itself.

Firstly, we're privileged to be able to say that we've been granted the use of this marvellous photo by Tony Blair's official photographer, who rang us up and asked us if we'd like to use it on the website. Of course, after we'd picked ourselves up off the floor, we said 'Yes!' So many thanks to Keith Taylor - the image below is by Keith Taylor and may not be reproduced without his permission.
Legal Deposit

• From 1610-1710-2013
• Unique arrangement of collaboration:
  – across two jurisdictions,
  – four countries,
  – Six libraries
• NPLD implementation began in 2013
• Four nodes (London, Boston Spa, Scotland, Wales)
• Core team in BL, NLW, NLS
• LD Implementation Group drawn from across the six libraries
• Sub-Groups: Content, Technical, Metadata, Security, Maps, Music, Emerging Formats etc
What is Legal Deposit?

• “Legal deposit helps to ensure that the nation’s published output, and thereby its intellectual record and future published heritage, is collected systematically, both in order to preserve the material for the use of future generations and to make it available for readers within the designated deposit libraries.” — British Library website

• Benefits:
  • Supporting research and culture
  • Supporting innovation and economic value creation

• Balance of responsibilities:
  • Publishers bear responsibility of depositing
  • Deposit libraries bear responsibility of preserving

• Balance of interests:
  • Rights holders: intellectual property business models
  • Public: national collection with continuity of access
What content is in scope?

- Web sites and pages
- XML/SGML e-journals
- PDF e-journals and monographs
- EPub e-books
- Born Digital Sheet Music
- Born Digital Mapping
- Emerging electronic formats
- Database publications
Legal Deposit

• Some numbers:
  – Via Portico
    • 2.3m journal articles to date
    • 90k added in April 2017
  – Direct from publishers
    • 530k to date
    • 14k added in April 2017
  – 135k ebooks to date
  – 5k added in April 2017
Institutional Research Outputs

- Research Data has become the new ‘hot-topic’ for librarianship and scholarly communications.
- Actually nothing new ...
- EPSRC mandate has provided a driver for institutions to move into this space.
- Range of supporting services
- Long-term Preservation not a key concern for academics or research funders at present.
- Obvious issue over re-usability and verifiability: new area of Research Ethics.
Dr Fraud and the strange case of the online journals

Anjana Ahuja

Professor Anna O Szust is exactly the kind of trophy academic that the University of Cambridge loves.
Deposit in the Oxford University Research Archive (ORA)

Oxford Researcher
- Article AAM (Author Accepted Manuscript)
- Journal Article (Quick deposit)
- Data
- Working/Discussion paper, report etc
- Conference Paper
- Other Research Item
- Bulk Upload

To deposit items in ORA you need your Oxford Single-Sign-On (WebAuth) username and password. For more information see IT Services website.

Oxford Research Student
- Thesis
- Journal article (Quick deposit)
- Conference paper
- Working/Discussion paper, report etc

To deposit items in ORA you need your Oxford Single-Sign-On (WebAuth) username and password. For more information see IT Services website.

Oxford Alumni
- Thesis

Alumni who would like to deposit their thesis in ORA and whose Oxford username and password have expired should contact ORA staff at ORA@bodleian.ox.ac.uk to obtain a username and password.

If you encounter problems trying to make your submission to ORA please contact ORA@bodleian.ox.ac.uk and we will arrange an alternative method for the upload of your work to ORA. Apologies for any inconvenience caused.
1. Identify key conclusions from the paper and a subset of data from the paper to be replicated.

2. Obtain the materials and methods from the key figures of the paper and annotate them to determine the holes in the protocols that would influence the success of replication.

3. Contact laboratories to get them to fill in the holes. This often involves searching through old research notebooks, connecting with study authors that have since moved on, and interfacing with collaborators.

4. Compile obtained information into organized registered report that is published at eLife and undergoes a review process, open to commentary from study authors.

5. Perform experiments, *in vitro* and *in vivo*, according to protocol from registered report.

6. Publish findings in an open access Replication Study.
Organisational Records

• Archives for institutions like universities and other large-scale administrative structures have relied on process to organise and appraise records for a long time.
• Electronic records management in place in some but not all organisations to support this process.
• Driver for preservation is mostly legal and evidential, partly historical.
• Some organisations see benefit in having historical material for advertising.
"Unless Europe can be saved...

civilisation is lost."

"We must come to them ... as physicians and surgeons at work to heal this wounded and festering continent, the mother of Christendom." Prof. Gilbert Murray, O.M. (one of our Hon. Trustees)

THE MAYOR OF VIENNA says his city has reached a peak of misery unknown even in the famine after the first world war.

50,000 SCHOOL CHILDREN have been medically examined in Vienna and 70% were suffering from malnutrition.

IN POLAND shoes are almost impossible to get. IN KIEL 9,100 children are entirely without shoes.

OXFORD RELIEF is entrusted to Friends Relief Service whose voluntary workers* act with care, economy and a personal touch regardless of frontiers, race, politics, class or creed. Many countries have benefited and now the peoples of Austria, Poland and the British Zone of Germany are getting most attention. F.R.S. personnel is less than 40% Quaker, 20% C. of E., many Free Church etc., and some Roman Catholics; united in service.

Gifts as below will be welcome

(IF no local body now collecting for Europe)—
MONEY FOR FOOD, MEDICAL SUPPLIES, Etc., for children, the aged, and special cases. (This does not reduce our rations.) Please make cheques payable to "Gradely Relief," c/o Barclays Old Bank, High Street, Oxford. Sir Alan Pim will acknowledge.
CLOTHING, FOOTWEAR, UNIFORMS, TOWELS, BEDDING, BLACKOUTS, MENDING MATERIALS, Etc. (including clean, mendable articles for repair by Displaced Persons badly needing work), to "Gradely," 42, Hythe Bridge Street, Oxford. Lady Franks will acknowledge (if address supplied).
BOOKS AND PERIODICALS (any language; for Reading Rooms) to "GRADELY," c/o C.O.B.S.R.A., 75 Victoria St., London S.W.1. (M.G.W., "Lancet," etc., 1939-47, also needed.)

OXFORD FOR FAMINE RELIEF COMMITTEE (Registered War Charities Act, 1940) FOR THE FRIENDLESS

N.B.—We cannot send clothing parcels to individuals.
Oxfam GB
FIRST WEBSITE
-1996.
COMMENDED IN PARLIAMENT.

THE OXFORD COMMITTEE FOR FAMINE RELIEF HAS MADE AN APPEAL FOR DONATIONS AND HOPE INDIVIDUALS WILL RESPOND TO THIS APPEAL.
OXFORD COMMITTEE FOR FAMINE RELIEF
CLOTHES MONEY
REFUGEE APPEAL
THIS WEEK

LOCAL APPEAL FOR REFUGEES
Bring clothing here today

PLEASE help us

DANCE
VICTORIA CROSS GALLERY
SATURDAY MAY 21st
IN BENEFIT OF REFUGEE FUND
OXFAM GIFT SHOP
467, Oxford Street

EVERY ARTICLE HAS BEEN A GIFT.
EVERY ONE YOU BUY HELPS A PERSON IN NEED

COME IN AND BROWSE

OXFAM Gift Shop
and Gift Reception
<table>
<thead>
<tr>
<th>TRACK no.</th>
<th>TRACK TITLE</th>
<th>TRACK TIME</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Spurs players</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-Ledley King</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+ Jermaine Defoe</td>
<td></td>
</tr>
</tbody>
</table>

**Photographer:** Lionel Hemming  
[tel: 07790 025 393]  
[Email: lhemming @hotmail.com]  
[Website: philips.com]
KATIE
ROBOTS
ZOT BALL
MARAWI!
SELECT + UPLOAD.
Social Media & Transactional Data

- Increasing cultural and commercial dominance of Facebook and other social media titans online.
- No-one is currently archiving Facebook (including Facebook).
- Twitter is archived at the Library of congress, but only for the period, 2006-2010. No access is currently provided.
How do we know we are succeeding?

• Standards:
  – ISO16363:2012
  – European Framework for Audit and Certification for Digital Repositories
  – Data Seal of Approval
Find all the necessary information to assess your digital repository and apply for the Data Seal of Approval.

—— News & Events ——

New Standards and Certification entity on the horizon »

—— Seals Acquired Around the World ——
How do we know we are succeeding?

• Standards:
  – ISO16363:2012
  – European Framework for Audit and Certification for Digital Repositories
  – Data Seal of Approval
    • Unified Requirements for Core Certification of Trustworthy Digital Repository
      – Content Preservation
      – Integrity and Authenticity
      – IPR and Regulatory constraints
      – Expert guidance
      – Technical infrastructure
      – Security
Knowing ourselves

• Polonsky Digital Preservation at Oxford and Cambridge
  – Work collaboratively as a distributed team.
  – Each institution has its own Project Team
  – Join Project Team
  – Collaborative with the Digital Preservation Coalition

• Aim to build capacity, knowledge and momentum in DP.
Knowing ourselves

• Staff survey at Bodleian shows 80% of staff regard digital preservation as being of equal value as physical preservation
• But we also found a lack of knowledge about DP terms and concepts.
• We will be adding basic DP to all staff induction courses
• As well as tailored training for specific groups of staff.
Knowing ourselves

• Bodleian Digital Image Audit
  – 129 million image files from Google Books project PLUS
  – 5.8 million in-house generated image files (on tape) OF WHICH
    • 2.4 million currently available online via a Bodleian website.
  – Found locations and management regimes of all digital image collections (WE HOPE)

• Spread across multiple and aging web applications and tape storage.
• Undergoing a process of rationalising web sites
• Reengineering workflows to ensure DP
• Writing detailed data management plans
Knowing ourselves

– Follow them at:

https://www.dpoc.ac.uk/about/blog/
Digital Preservation at Oxford and Cambridge
A collaborative research project to evaluate and provide sustainable recommendations for our digital preservation programmes

About the Project
The Polonsky Digital Preservation Project is a collaboration between Bodleian Libraries (Oxford) and Cambridge University Library (Cambridge) with funding from the Polonsky Foundation.
Competition and the danger of complacency

- Dropbox
- Apple Time Machine + External Hard-drive
- iCloud
- Perception that storage=Preservation
Big Ideas

At the DPC's Planning Day at the British Library on 3rd April, we asked attendees to come up with some 'Big Ideas' for the coalition's next 5 years. Now it's time for the rest of the membership (our Dragon's Den) to vote on and choose these ideas before we add them to our next Strategic Plan, due to be launched at the end of this year.

We are building:

- A political and institutional climate responsive to the need for digital preservation
- Competent and responsive workforces ready to address the challenges of digital preservation
- Better tools, smarter processes and enhanced capacity in digital preservation
- Closer and more productive collaboration within and beyond the Coalition

Digital Preservation Handbook

The Handbook aims to identify good practice in creating, managing and preserving digital materials. By providing a strategic overview of the key issues, discussion and guidance on strategies and activities, and pointers to key projects and reports, the Handbook aims to provide guidance for institutions and individuals and a range of tools to help them identify and take appropriate actions.
and new libraries…
and new...
Martin Joseph Routh

Dr. Martin Joseph Routh (18 September 1755 – 22 December 1854) was President of Magdalen College from 28 April 1792 until his death.

This People article is a stub. You can help Wikiquote by expanding it.

Quotes [edit]

- I think sir, since you care for the advice of an old man, sir, you will find it a very good practice, always to verify your references, sir!
  - Advice given to Dean John William Burgon, (29 November 1847), in response to the question: “Every studious man, in the course of a long and thoughtful life, has had occasion to experience the special value of some one axiom or precept. Would you mind giving me the benefit of such a word of advice?”; quoted in Lives of twelve good men, by John William Burgon, 1888, vol. 1 p. 73.

- It will lead to nothing, I fear, sir
  - Remarked to William Palmer, on the eve of his 1840-1 journey to Russia to improve Anglican-Orthodox relations. The trip reaped little success; quoted in Notes of a Visit to the Russian Church in the Years 1840, 1841, by William Palmer. 1882. p. 10.

- But I know, my friends, that you may object to me what St. Irenaeus says.
  - Said to his congregation after an exposition of his own theology; quoted in Essays and addresses, Herbert McLachlan, Manchester University Press, 1954, p. 354

- Wait, sir, until I am gone!
  - Standard reply to people proposing changes in the running of Oxford University; quoted in Colin Gordon, Beyond the Looking Glass (1962), p. 37

External links [edit]

Wikipedia has an article about: Martin Joseph Routh

Categories: People stubs | Educators | English people | 1850s deaths
EU referendum: Vote Leave campaign gets underway in Cornwall

With six weeks to go until the EU referendum, Vote Leave unveil their campaign and battle bus in Cornwall.

Vote Leave battle bus unveiled
Richard Ovenden
Bodley’s Librarian

@richove
richard.ovenden@bodleian.ox.ac.uk
+44 1865 277158