



# WARCnet

## *Special Reports*

Scholarly Use of Web Archives Across  
Ireland: The Past, Present & Future(s)

# Scholarly Use of Web Archives Across Ireland: The Past, Present & Future(s)

*Sharon Healy (Maynooth University)*  
*Helena Byrne (British Library)*

schealy.ire@gmail.ie

helena.byrne@bl.uk



WARCnet Special Report  
Aarhus, Denmark 2023

Sharon Healy, Helena Byrne: *Scholarly Use of Web Archives Across Ireland: The Past, Present & Future(s)*  
© The authors, 2023

Published by the research network  
WARCnet, Aarhus, 2023.

Editors of WARCnet Special Reports:  
Niels Brügger, Jane Winters, Valérie  
Schafer, Kees Teszelszky, Peter  
Webster, Michael Kurzmeier.

Cover design: Julie Brøndum, Kamilla  
Rosenberg, Emma Lund Nielsen, Thea  
Laugesen  
ISBN: 978-87-94108-17-1

WARCnet  
Department of Media and Journalism  
Studies  
School of Communication and Culture  
Aarhus University  
Helsingforsgade 14  
8200 Aarhus N  
Denmark  
warcnet.eu

The WARCnet network is funded by the  
Independent Research Fund Denmark |  
Humanities (grant no 9055-00005B).





## WARCnet Papers

Niels Brügger: *Welcome to WARCnet* (May 2020)

Ian Milligan: *You shouldn't Need to be a Web Historian to Use Web Archives* (Aug 2020)

Valérie Schafer and Ben Els: *Exploring special web archive collections related to COVID-19: The case of the BnL* (Sep 2020)

Niels Brügger, Anders Klindt Myrvoll, Sabine Schostag, and Stephen Hunt: *Exploring special web archive collections related to COVID-19: The case Netarkivet* (Oct 2020)

Friedel Geeraert and Nicola Bingham: *Exploring special web archives collections related to COVID-19: The case of the UK Web Archive* (Nov 2020)

Friedel Geeraert and Barbara Signori: *Exploring special web archives collections related to COVID-19: The case of the Swiss National Library* (Nov 2020)

Matthew S. Weber: *Web Archives: A Critical Method for the Future of Digital Research* (Nov 2020)

Niels Brügger: *The WARCnet network: The first year* (Jan 2021)

Susan Aasman, Nicola Bingham, Niels Brügger, Karin de Wild, Sophie Gebeil and Valérie Schafer: *Chicken and Egg: Reporting from a Datathon Exploring Datasets of the COVID-19 Special Collections* (Dec 2021)

Olga Holownia, Friedel Geeraert, Abbie Grotke, Jennifer Harbster and Gulnar Nagashybayeva: *Exploring special web archives collections related to COVID-19: The case of the Library of Congress* (Feb 2022)

Niels Brügger: *The WARCnet network: The second year* (Dec 2022)

Michael Kurzmeier: *Using a national web archive for the study of web defacements? A case-study approach* (Aug 2023)

Helle Strandgaard Jensen: *Any Teletubbies Caught in the Web?* (Aug 2023)

Niels Brügger: *The WARCnet network: The third year* (Aug 2023)

Valérie Schafer, Jérôme Thièvre and Boris Blanckemane: *Exploring special web archives collections related to COVID-19: The case of INA* (Aug 2020)

Niels Brügger, Valérie Schafer, Jane Winters (Eds.): *Perspectives on web archive studies: Taking stock, new ideas, next steps* (Sep 2020)

Friedel Geeraert and Márton Németh: *Exploring special web archives collections related to COVID-19: The case of the National Széchényi Library in Hungary* (Oct 2020)

Friedel Geeraert and Nicola Bingham: *Exploring special web archives collections related to COVID-19: The case of the IIPC Collaborative collection* (Nov 2020)

Caroline Nyvang and Kristjana Mjöll Jónsdóttir Hjörvar: *Exploring special web archives collections related to COVID-19: The Case of the Icelandic web archive* (Nov 2020)

Sophie Gebeil, Valérie Schafer, David Benoist, Alexandre Faye and Pascal Tanesie: *Exploring special web archive collections related to COVID-19: The case of the French National Library (BnF)* (Dec 2020)

Karin de Wild, Ismini Kyritsis, Kees Teszelszky and Peter de Bode: *Exploring special web archive collections related to COVID-19: The Dutch Web archive (KB)* (Nov 2021)

Michael Kurzmeier, Joanna Finegan and Maria Ryan: *Exploring special web archives collections related to COVID-19: The National Library of Ireland* (Feb 2022)

Emily Maemura: *Towards an Infrastructural Description of Archived Web Data* (May 2022)

Olga Holownia, Friedel Geeraert and Paul Koerbin: *Exploring special web archives collections related to COVID-19: The case of the National Library of Australia* (Dec 2022)

Helena Byrne, Beatrice Cannelli, Carmen Noguera, Michael Kurzmeier, Karin de Wild: *Looking ahead: after web (archives)?* (Aug 2023)

Friedel Geeraert, Marie Haškovcová, Luboš Svoboda and Markéta Hrdličková: *Exploring special web archives collections related to COVID-19: The case of the National Library of the Czech Republic* (Aug 2023)

## WARCnet Special Reports

Sharon Healy, Helena Byrne, Katharina Schmid, Nicola Bingham, Olga Holownia, Michael Kurzmeier, Robert Jansma: *Skills, Tools, and Knowledge Ecologies in Web Archive Research* (Aug 2022)

Sharon Healy, Helena Byrne: *Scholarly Use of Web Archives Across Ireland: The Past, Present & Future(s)* (Aug 2023)

## Author Information

**Sharon Healy** is an independent researcher and archivist with strong interests in internet/web history, RDM workflows for digital data, and the preservation of digitised, born digital, and reborn digital heritage. Sharon holds a BA (Hons) in Cultural Studies, an MA in Digital Humanities, a PG Diploma in Historical Archives, and a PhD in Digital Humanities & Archives. This study forms part of the doctoral research of Sharon Healy at Maynooth University, which was kindly funded by the Irish Research Council under grant number [GOIPG/2018/1543].

**Helena Byrne** is the Curator of Web Archives at the British Library. She was the Lead Curator on the IIPC Content Development Group 2022, 2018 and 2016 Olympic and Paralympic collections. Helena completed a Master's in Library and Information Studies at University College Dublin, Ireland in 2015. Previously she worked as an English language teacher in Turkey, South Korea, and Ireland. Helena is also an independent researcher that focuses on the history of women's football in Ireland, and the fragility of digital gender histories. Her previous publications cover both web archives and sports history.

## Acknowledgements

We have many organisations and individuals to thank for their support, assistance, and encouragement throughout the duration of this study. First, we would like to thank the WARCnet Steering Group for organising network meetings and activities, which provided the capacity for members to develop projects such as this one. Next, we would like to thank Maynooth University and the British Library for providing back bone support to the authors. Also, we must sincerely thank the respondents of the survey study for their participation, for without them, the full range of this report would not have been possible. We express our gratitude and appreciation to the external reviewers in giving up their time to read the final drafts of this report and providing us with valuable feedback, including Nicola Bingham from the British Library, John Sterne from TechArchives (Ireland), and staff members from the National Library of Ireland, the Library of Trinity College Dublin, and the Public Record Office of Northern Ireland. Finally, we would like to acknowledge that this study forms part of the doctoral research of Sharon Healy which was kindly funded by the Irish Research Council [GOIPG/2018/1543], under the supervision of Dr. Joseph Timoney, Dr. Martin Maguire, Dr. Juan-José Boté-Vericad and Mr. Thomas Lysaght.

## Accessibility

As part of our commitment to accessibility, we have tried to ensure that the URLs provided in this report are (i) captured in a web archive close to the time of access on the live web or (ii) saved in a web archive close to the time of access on the live web. In case of future link rot, we have documented which archive the URL may be found in the report Bibliography, e.g. [URL Memento: Wayback Machine]. An accompanying dataset of bibliographic files (e.g., BibTex, CSL JSON, CSV, etc.) will also become available to download through our project files, available in Open Science Framework (<https://osf.io/crfnw/>). Regarding paywall journal articles, we have attempted to provide a DOI, when available, and in the case of open access journal articles, we have further attempted to capture the source URL in a web archive. To further assist with accessibility, we utilise the Arial/Calibri fonts, and apply [alt text] for all images contained in this document. Should a reader need to access this document in some other form which would provide better accessibility, please contact the authors.

## Glossary of Terms

In providing a glossary we direct the reader to a glossary published as part of the WARCnet WARST project, titled 'Towards a Glossary for Web Archive Research: Version 1.0'. An interactive glossary resource is also being developed online in Zotero Groups.

Healy, S., Byrne, H., Schmid, K., Boté-Vericad, J.-J., Floody, L. (2023). Towards a Glossary for Web Archive Research: Version 1.0. *WARCnet Special Reports*. Aarhus, Denmark: WARCnet.

Healy, S., Byrne, H., Schmid, K., Boté-Vericad, J.-J., Floody, L. (2021+). Zotero Groups - Towards a Glossary for Web Archive Research. Zotero, [https://www.zotero.org/groups/4380600/towards\\_a\\_glossary\\_for\\_web\\_archive\\_research](https://www.zotero.org/groups/4380600/towards_a_glossary_for_web_archive_research).

## Abbreviations

AADDA	Analytical Access to the Domain Dark Archive
API	Application Programming Interface
ARC_IA	Internet Archive ARC file format
BnF	Bibliothèque nationale de France
BUDDAH	Big UK Domain Data for the Arts and Humanities project
BBS	bulletin board system
ccTLD	country code Top Level Domain
CD-ROM	Compact Disc-Read Only Memory
COIPLPA	Copyright and Other Intellectual Property Law Provisions Act 2019 [Ireland]
CRC	Copyright Review Committee [Ireland]
CRRA	Copyright and Related Rights Act, 2000 [Ireland]
DH	Digital Humanities
DMP	Data Management Plans
DVD	Digital Video Disc
EU	European Union
EC	European Communities
EEC	European Economic Community
GDPR	General Data Protection Regulation
HTML	HyperText Markup Language
ICPPA	Industrial and Commercial Property (Protection) Act, 1927 [Ireland]
IIPC	International Internet Preservation Consortium
IFLA	International Federation of Library Associations
IMF	Internet Memory Foundation
IRC	Irish Research Council
IT	Information Technology



JISC	Joint Information Systems Committee [United Kingdom]
NAI	National Archives of Ireland
NI	Northern Ireland
NLI	National Library of Ireland
NDSA	National Digital Stewardship Alliance [United States]
NPLD	The Legal Deposit Libraries (Non-Print Works) Regulations 2013 [United Kingdom]
ODU WS-DL	Old Dominion University, Web Science and Digital Libraries Research Group [United States]
PANDORA	Preserving and Accessing Networked Documentary Resources of Australia
PDF	Portable Document Format
PNG	Portable Network Graphics
PROI	Public Record Office of Ireland
PRONI	Public Record Office of Northern Ireland
PWID URI	Uniform Resource Identifier for Persistent Web IDentifiers
ROI	Republic of Ireland
RESAW	Research Infrastructure for the Study of Archived Web Materials
SPO	State Paper Office
TCD	Trinity College Dublin [Ireland]
TCP/IP	Transmission Control Protocol/Internet Protocol
TEI	Text Encoding Initiative
UK	United Kingdom
UKWA	UK Web Archive
UKWAC	UK Web Archiving Consortium
UNESCO	United Nations Educational, Scientific and Cultural Organization
UN	United Nations
URI	Uniform Resource Identifier
URL	Uniform Resource Locator

US	United States
WARC	Web ARChive file format
WARCnet	Web ARChive studies network researching web domains and events
WARST	Web Archives - Researcher Skills & Tools Survey
XML	EXtensible Markup Language

# List of Contents

EXECUTIVE SUMMARY .....	i
Abstract.....	i
Key Findings and Insights.....	ii
1. INTRODUCTION .....	1
1.1 Background.....	2
1.2 Purpose of the Study .....	9
1.3 Document Outline .....	10
2. RECOGNISING THE PROBLEMS .....	12
2.1 Ephemeral Nature of the Web .....	12
2.2 From Print to Non-Print .....	23
2.3 Web Archives and Scholarly Engagement .....	28
2.3.1 Challenges for scholarly engagement with web archives.....	28
2.3.2 Scholarly use of web archives .....	39
2.4 Web Archive User Studies.....	43
3. AVAILABILITY AND ACCESSIBILITY OF IRISH BASED WEB ARCHIVES.....	49
3.1 Introduction.....	49
3.2 Preservation of Irish Records and Publications, pre-Digital .....	52
3.3 NI Web Space .....	57
3.3.1 PRONI web archive .....	57
3.3.2 UK web archive .....	59
3.3.3 NI web space in brief .....	66
3.4 ROI Web Space .....	67
3.4.1 NLI web archive .....	68
3.4.2 Current status of the NLI web archive .....	85
3.4.3 Web archives in the Irish media .....	86
3.4.4 ROI web space in brief .....	88
4. AWARENESS OF, AND ENGAGEMENT WITH, WEB ARCHIVES IN IRISH ACADEMIC INSTITUTIONS .....	89
4.1 Introduction.....	89
4.2 Related Literature .....	91
4.2.1 Web archive user studies .....	91
4.2.2 Use of web archives for Irish based research .....	91

4.3 Methodology .....	94
4.3.1 Survey software .....	95
4.3.2 Survey recruitment .....	95
4.3.3 Survey design and questions .....	96
4.3.4 Survey responses .....	98
4.3.5 Survey limitations .....	99
4.4 Results and Analysis .....	100
4.4.1 Demographics .....	100
4.4.2 Engagement with online digital-based resources .....	105
4.4.3 Awareness of the existence of web archives .....	106
4.4.4 Engagement with web archives for personal and research interests .....	109
4.4.5 Non-users of web archives for research .....	109
4.4.6 User Engagement with web archives .....	114
4.4.7 Perceived value and importance of web archives .....	121
4.4.8 Perceived challenges for the use of archived web content for studies or research in the future .....	124
4.5 Discussion .....	130
4.5.1 Current level of awareness for the existence of web archives .....	130
4.5.2 Terminology .....	132
4.5.3 Reasons for a lack of engagement with web archives for research .....	132
4.5.4 Likelihood of a non-user using a web archive for research, after becoming aware of its existence .....	134
4.5.5 Challenges perceived by scholars for the future use of archived web content .....	134
4.5.6 Users of web archives in Irish academic institutions .....	135
4.5.7 Perceived importance of archiving websites based on specific topics .....	137
4.5.8 Perceived value of web archives .....	138
4.6 Summary .....	138
5. CONCLUSIONS .....	139
5.1 Key Findings and Insights .....	139
5.1.1 Main causes for the loss of digital heritage .....	139
5.1.2 Availability and accessibility of web archives based on the island of Ireland for conducting Irish based research .....	142
5.1.3 Scholarly engagement and non-engagement with web archives .....	144
5.2 Final Thoughts and Future Work .....	151
BIBLIOGRAPHY .....	153
APPENDICES .....	198

## List of figures

Figure 1.1: Web Archiving Life Cycle Model (Bragg & Hanna, 2013) which is inclusive of appraisal, selection, capture, storage, quality assurance, preservation and maintenance, replay/playback, access, use and reuse .....	3
Figure 2.1: The changing nature of the Government of Ireland website captured in the Wayback Machine from 1996 to 2008 (www.irlgov.ie), and 2008 to 2011 (www.gov.ie) .....	18
Figure 2.2: Screenshot of the IIPC about/index page, captured in 2004 in the Wayback Machine (Timestamp: 2004-06-03 01:41:15) .....	22
Figure 2.3: Screenshot of the homepage for the web version of the interactive CD-ROM ‘Safeguarding our documentary heritage’ created by UNESCO and IFLA-PAC which provides interactive guidelines on the preservation of archival materials including multimedia formats, captured in the Wayback Machine (Timestamp: 2001-04-29 03:33:55). .....	24
Figure 3.1: Screenshot of the interface of the 3D Virtual Record Treasury of Ireland ( <a href="https://vrtour.virtualtreasury.ie">https://vrtour.virtualtreasury.ie</a> ), taken on 2022-10-18.....	55
Figure 3.2: Screenshot of the interface of the PRONI Web Archive, taken on 2022-08-24 .....	58
Figure 3.3: Screenshot of the interface of the PRONI Web Archive, showing descriptive metadata entries, taken on 2022-09-26.....	59
Figure 3.4: Screenshot of UKWA web page for Topics and Themes, taken on 2022-10-16.....	66
Figure 3.5: Screenshot of NLI Web Archive user interface on the Internet Memory Foundation platform, taken in June 2015 (personal archive) .....	69
Figure 3.6: Screenshot of NLI Web Archive interface on the Archive-It platform showing a total of 3,114 websites in their collections, taken on 2023-01-16.....	70
Figure 3.7: Screenshot of NLI Web Archive interface on the Archive-It platform showing a total of 75 collections, taken on 2023-01-16.....	70
Figure 3.8: Screenshot of Section 108 in the Copyright and Other Intellectual Property Law Provisions Act 2019, taken on 2022-10-07 .....	71
Figure 3.9: The changing nature of ROI Government Department websites and URLs is revealed by examining the first captures of their websites in the Wayback Machine (see Table 3.1) .....	76
Figure 3.10: Screenshot of NLI Web Archive interface, showing multiple captures of the Sinn Féin website from 2011 to 2022 (www.sinnfein.ie), taken on 2022-10-06 .....	87
Figure 4.1: Position of users (n=59) under the representations of educators, researchers, and students, in line with total responses (N=239) .....	104
Figure 4.2: Representation of participant engagement with web archives for personal interests and research.....	109

Figure 4.3: Representation for the likelihood of future engagement by a non-user (n=180) with a dark web archive.....	114
Figure 4.4: Representation of general reasons for using a web archive (n=59).....	116
Figure 4.5: Representation for the likelihood of future engagement by users (n=59) with a dark web archive.....	120

## List of tables

Table 2.1: Useful self-archiving web services for researchers and other users .....	38
Table 3.1: Renaming of Department of Jobs, Enterprise and Innovation before and after formulation (Sources: Wikipedia, 2005+).....	76
Table 4.1: Breakdown of recruitment emails sent per university .....	96
Table 4.2: Representation of participant responses for nationality (N=239), with a comparison of nationality representations for users and non-users .....	101
Table 4.3: Representation of participant responses for age (N=239), with a comparison of age representations for users and non-users.....	102
Table 4.4: Representation of participant responses for gender (N=239), with a comparison of gender representations for users and non-users .....	102
Table 4.5: Representation of participant responses for position (N=239), with a comparison of position representations for users and non-users.....	103
Table 4.6: Representation of participant responses for discipline category (N=239), with a comparison of discipline representations for users and non-users .....	104
Table 4.7: Representation of participant responses for engagement with other online/digital resources .....	106
Table 4.8: Representation of a comparison of discipline categories of respondents who indicated awareness of the online public NLI Web Archive .....	107
Table 4.9: Representation of participant responses (N=239) for awareness of other online public web archives .....	108
Table 4.10: Representation of non-user respondent (n=180) reasons for not using an online web archive for their studies/research .....	110
Table 4.11: Representation of discipline categories for non-user respondents who indicated a lack of research engagement with online web archives due to a lack of awareness (n=141), in line with the total number of user and non-user participants who identified with that discipline category .....	111

Table 4.12: Representation of non-user responses (n=180) for the likelihood of future engagement with online public web archives.....	112
Table 4.13: Representation for the probability that awareness increases likelihood of engagement with online public web archives for non-users (n=180) who were unaware of the existence of online public web archives.....	113
Table 4.14: Representation of discipline categories for user respondents (n=59) .....	115
Table 4.15: Representation of user participant reasons for using web archived content for their studies/research .....	117
Table 4.16: Representation of user respondent reasons for using web archives for study or research (n=59) .....	118
Table 4.17: Representation for the use of online public web archive by user respondents .....	119
Table 4.18: Representation of participant responses (N=239) for their perceived value of web archives .....	121
Table 4.19: Representation of participant responses (N=239) on the importance of archiving websites/blogs based on topics .....	123
Table 4.20: Representation of participant responses (N=239) on the importance of web archives for current, medium, or long-term future research .....	124
Table 4.21: Thematic representation of participant responses on their perceived challenges for the future use of archived web content in their field of research (n=50) .....	125
Table 5.22: Combined data from Section 3.6 for user participant (n=59) reasons for using archived web content for their studies or research .....	137
Table D.1: Breakdown for position and discipline categories of respondents who indicated that they use the online public NLI Web Archive for their studies/research (=23).....	210
Table E.1: Discipline categories for respondents (N=239) who indicated 'Yes' on the importance of web archives for current, medium, or long-term future.....	211

# EXECUTIVE SUMMARY

## Abstract

This report is the product of a collaborative study by Sharon Healy (Maynooth University) and Helena Byrne (British Library). The study incorporated a review of resources and literature, informal dialogues with heritage colleagues and the use of an online survey. The study sought to (i) examine the causes for the loss of digital heritage and how this relates to Ireland, (ii) offer an overview of the landscape of web archives based across Ireland, and their availability, and accessibility as resources for Irish based research, and (iii) provide some insight into the awareness of, and engagement with, web archives in Irish third-level academic institutions.

First, the report examines some of the main causes for the loss of digital heritage and how this relates to Ireland. In doing so, it explores the relationship between legal deposit legislation and the preservation of national heritage and observes how web archiving is a necessary activity for the preservation of digital heritage. Thereafter, the report investigates scholarly engagement with web archives, through a review of user case studies and web archive user studies, and discusses the challenges experienced by the web archive user community, as well as the opportunities for using the web archives for research purposes. Next, the report assesses the availability and accessibility of web archives based on the island of Ireland, and their usefulness as resources for Irish based research with a focus on the PRONI Web Archive, the UK Web Archive and the NLI Web Archive. Through a synthesis of legal deposit history, web history, and political debates and inertia, the report illustrates the need for the continual evaluation of legal deposit legislation in line with the fragility of born digital heritage and the technological advances in publishing and communication technologies. To end, using an online survey, the report investigates awareness of web archives, and engagement or non-engagement with web archives in Irish academic institutions.

This report will be of benefit not only for web archive users but also the wider web archiving community as many of the challenges faced by Irish based web archiving initiatives and Irish based researchers will not be unique to Ireland.

In the following sections we offer an overview of the key findings and insights, through an examination of the main causes for the loss of digital heritage, the availability and accessibility of web archives based on the island of Ireland for conducting Irish based research, and the scholarly engagement and non-engagement with web archives in Irish academic institutions.



## Key Findings and Insights

The focus of this report, for the most part, is on institutional web archiving and curation through web crawling, and the use of institutional web archives for research or other purposes. In addition, this report refers to Irish digital heritage in the context of the digital heritage of the island of Ireland, and when required, it refers to the digital heritage of Northern Ireland (NI) or the Republic of Ireland (ROI) to distinguish between the two jurisdictions.

### Main causes for the loss of digital heritage

Within a few years of the web becoming established as a new medium for publishing and sharing information, national libraries and cultural heritage organisations became concerned about the ephemeral nature of the web, and instigated preservational strategies for the capture and preservation of digital heritage on the web through web archiving. These concerns were further substantiated by studies which examine link rot, reference rot, and web content change over time. There are several reasons put forward as to why web content moves, changes, or gets deleted. These include software and system upgrading, changes in filing systems, the re-arrangement of web content, the relocation of servers, a lack of funding or interest to maintain websites, and simply a lack of foresight by web publishers. Moreover, concerns for the loss of digital heritage on the web stemmed from wider concerns about the appraisal, storage and long-term preservation of electronic information, multimedia and born digital materials in general. These wider concerns have been around since before the web was invented, with the web just becoming another media carrier to worry about. Indeed, for UNESCO (2003) “the disappearance of heritage in whatever form constitutes an impoverishment of the heritage of all nations”, and digital heritage should not be an exception.

Some of the factors which contribute to the loss of digital heritage to posterity include technological obsolescence of hardware and software, media deterioration, availability of resources, and inadequate legislation (UNESCO, 2003; Waters & Garrett, 1996; Besser, 2000). Furthermore, the loss of digital heritage has often gone unnoticed by societies and nations because “Attitudinal change has fallen behind technological change” and consequently, the economic, social, intellectual, and cultural value or potential value of the heritage is not realised (UNESCO, 2003). For Lyman (2002), societies have lost important parts of their cultural heritage in the past because it was not archived or preserved due to cultural, technical, economic, and legal problems. The cultural problem is due to the inability of past generations to recognise its importance and historic value, while the technical problem is due

to a lack of foresight and technical ingenuity to ensure continuity for preservation, storage, and maintenance (Lyman, 2002). Lyman (2002) posits how the economic problem stems from the failure to find a business model to support the archiving of new media formats, while the legal problem stems from the failure to create legislation which protects copyright while at the same time allows for archival preservation. These problems equally apply to the loss of digital heritage on the web.

The evolving nature of publishing over the past 50 years has become problematic for legal deposit legislation which was fundamentally print-centric. For hundreds of years the concept of legal deposit served as a system to compile and preserve a collection of a country's publications outputs, providing a significant contribution to national cultural heritage. As a result, several countries began to amend their copyright and legal deposit legislation from the 1990s to accommodate the deposit of non-print materials and media formats (e.g., microfilm, CD-ROM, DVD etc.) and for born digital materials, inclusive of the web archiving of a country's national web domain, as a matter of routine. On the other hand, for many countries legal deposit legislation is still outdated with regard to emerging publishing technologies and the advances in internet and web technologies. For example, the ROI has trailed behind Canada, New Zealand, and much of Europe (Conul, 2012, p. 14). Consequently, one of the major causes for the loss of Irish digital heritage to posterity, is the failure of successive ROI governments to negotiate copyright and legal deposit legislation in line with advances in publishing and communications technologies, and due to the current deficiencies of ROI copyright and legal deposit legislation to include the routine web archiving of the Irish national domain as part of a national legal deposit scheme.

In the UK, legal deposit legislation was reformed in 2003 in such a way that established the basis for the selective web archiving programme undertaken by the UK Web Archiving Consortium, which included the capture and preservation of websites from the NI web space. The 2003 non-Print Legal Deposit Act mandated the responsible minister with the powers to bring in regulations for digital collecting, including websites, under legal deposit which could be enacted at the appropriate time in the future. After the practical details had been worked out, the regulations were eventually applied in 2013. This meant that, from 2003 - 2013, the UK Legal Deposit Libraries (LDLs) selectively archived websites under existing copyright law. In the approximately ten years from 2003, the LDLs contributed to the discussion about whether digital collecting needed legislation or whether it could be carried out under voluntary deposit. The conclusion was that seeking permission to archive from website publishers was not feasible and the regulations were necessary. The legislation was therefore updated in 2013 to allow for an annual web crawl of the UK web estate, including NI,

undertaken by the UK Web Archive, a partnership of the six UK LDLs. In addition, the PRONI Web Archive commenced a selective web archiving initiative in 2010 to capture and preserve websites of NI government departments, local councils, public sector organisations and websites which have social, cultural, political, religious, or economic significance for the preservation of NI heritage. However, prior to 2013, the UK/NI web space was not systematically captured as part of legal deposit, and therefore much of the earlier NI webspace will have disappeared or changed drastically (Jackson, 2015a). To salvage some of the UK web estate prior to 2013, the Joint Information Systems Committee (JISC) acquired a dataset from the Internet Archive which included all .uk websites in their web archive collections that were crawled from 1996-2013 (UK Web Archive, n.d., JISC UK Web Domain Dataset).

The ROI also trails behind in terms of the preservation of governmental digital heritage. There have been continual warnings by the National Archives of Ireland (NAI) to the ROI government, since at least 1997, regarding the loss of digital heritage due to the lack of a “comprehensive formal records, management policy for State” and the “Loss of electronic records and archives or access to them, due to degeneration of storage media and/or redundancy of operating systems” (Reports of the Director of the National Archives of Ireland, 2014-2020). Regrettably, over twenty years since the problem was identified, the Irish government has still not come to terms with the preservation of electronic records, nor does it seem to have a formal policy for record keeping in any electronic format. In the meantime, content on the Irish government website(s) has changed and disappeared over the past decades, and Irish government department websites have been particularly vulnerable to link rot, and changes in website content.

## Availability and accessibility of web archives based on the island of Ireland for conducting Irish based research

Through an analysis of the availability and accessibility of web archiving initiatives based on the island of Ireland, and their usefulness for conducting Irish based research, the report offers insights which may be useful when it comes to assessing support and incentive mechanisms which will foster developments in the use of web archives for Irish based research and encourage engagement for a diverse range of end users (e.g., public administrators, journalists, legal professionals, web designers, local historians etc.). While acknowledging that Irish web heritage can be found in various international web archives, attention was given to web archiving initiatives which have a specific mandate to capture a wide range of Irish web heritage as part of their collection development strategies. Therefore, the focus was on the PRONI Web Archive, the NLI Web Archive, and the UK Web Archive,

which is accessible onsite in the Library of Trinity College Dublin (TCD). The report observes the efforts of these initiatives for the collection and preservation of digital heritage from the web spaces of NI and ROI, and offers an overview of their historical backgrounds, inclusive of how copyright and legal deposit has influenced their collecting activities.

In the case of the digital heritage of NI, the findings suggest that while there are resource and legislative limitations, there are nonetheless concrete efforts being made to provide a balanced approach towards the collection and preservation of the NI web space. First, the UK Web Archive captures and preserves websites from the NI web space, through a selective collection approach and through an annual domain crawl of the NI web space as part of legal deposit, which is accessible onsite in a UK legal deposit library, inclusive of the Library of TCD in Dublin. Second, NI digital heritage is preserved through a two-fold approach by PRONI to provide a publicly accessible selective web archive collection, through (i) the collection of websites of government, public bodies etc., with notifications of the intent to collect, and provisions of a takedown policy, and (ii) a permissions-based approach for privately funded websites. And third, the NI web space is preserved through a collaborative effort by the PRONI Web Archive and the UK Web archive for the development of accessible curated collections. While there is a wide range of topics within the collections of the UK Web Archive and the PRONI web archive which would be useful for conducting Irish based research, access to the collections differ. PRONI Web Archive is open access, and the UK Web Archive is a mix of both open access and onsite access. However, as discussed, onsite access presents challenges for researchers due to the restrictive nature of the access protocols in the current UK legal deposit legislation, which is outdated in line with advances in publishing and communications technologies, and current trends in digital user expectations and information seeking behaviours (Gooding et al., 2019).

Regarding the ROI, the National Library of Ireland (NLI) began a small-scale selective web archiving initiative in 2011, to include a wide range of topics which would be useful for conducting Irish based research. In addition, the NLI conducted two web domain crawls in 2007 and 2017, which are currently inaccessible to researchers or the public due to legislative matters. While the NLI is a legal deposit library, digital legal deposit legislation was not enacted in Ireland at the time the domain crawls were conducted. Moreover, while digital legal deposit legislation came into force in December 2019 through the Copyright and Other Intellectual Property Law Provisions Act 2019 (hereafter, COIPLPA, 2019), it did not include a clause for crawling the Irish national web domain. However, COIPLPA (2019) contains a clause to “bring forward a report on the feasibility of establishing a digital legal deposit scheme to serve as a web archive for .ie domain contents and advise on steps taken towards that goal”

within twelve months of the Act coming into force in December 2019. However, as of May 2023, a feasibility report has yet to be produced. Delays for this include an inertia related to political debates on the archiving of the Irish national web domain as part of legal deposit legislation in line with other countries in Europe.

Traces of the ROI web estate in other web archives are very shallow. Therefore, it will be impossible to retrospectively recreate the ROI web space, although it is imperative to salvage as much as possible. As it stands, the ROI is already “impoverished” (UNESCO, 2003) due to mass losses of digital heritage on the web for the decades of the 1990s, 2000s, and 2010s. It now looks like this will continue well into the 2020s, before the necessary measures are put in place for the collection and preservation of the web space of the twenty-six counties of the ROI in line with the collection and preservation of the web space of the six counties of NI.

Therefore, the report stresses that urgent action is required for an emergency change in ROI legislation to allow for the collection and preservation of the ROI web estate in the interim, while the feasibility report continues to be undertaken to advise on the necessary requirements to update the legislation, and to establish a national web domain archive through “a process of negotiation among interested parties” (Lyman, 2002). Furthermore, such negotiations should be inclusive of representatives from the education and teaching sectors, and end users such as academics, public administrators, journalists, legal professionals, and web designers, as well as information professionals who have experienced the transition from small-scale selective web archiving to large-scale domain web archiving, and information professionals who are experienced in working with Irish based information ecosystems, and thus, minimise the challenges from the start. There is also a need to assess the demarcation of the Irish national web domain, as using the .ie ccTLD is not an adequate marker for the representation of Irish digital heritage on the web. Finally, the report underscores how born digital content is more fragile than print material, and publishing and communications technologies are constantly changing. Thus, any legislation implemented will have to be reviewed on a regular basis in order to keep up with the changes in technology and current trends in digital user expectations and information seeking behaviours (Gooding et al., 2019). We further noted how the formation of a Copyright Council of Ireland, as suggested in the Modernising Copyright report (2013), could be tasked with monitoring legal deposit legislation in line with the fragility of born digital heritage and the technological advances in publishing and communication technologies.

## Scholarly engagement and non-engagement with web archives

In the next sections we sum up the challenges and opportunities for using web archives as resources for research and provide an overview of the findings on scholarly awareness, engagement, and non-engagement with web archives in Irish academic institutions.

### Challenges for scholarly engagement with web archives

There are several reasons put forward for the lack of scholarly engagement with web archives. Obvious reasons include a lack of awareness, or simply because some academic disciplines have no need to rely on such sources (Jatowt, 2008; Riley & Crookston, 2015; Winters, 2017; Costea, 2018). It can also be argued that a lack of dialogue or collaboration between the creators of web archives, and end users (or even potential end users) has had some effect on engagement with web archives for research purposes (Dougherty et al., 2010; Hockx-Yu, 2014; Schroeder & Brügger, 2017; Gooding et al., 2021). On a more positive note, collaboration between web archive creators and end user researchers has been improving over the past decade (Schroeder & Brügger, 2017; Webster 2017b; Maemura, 2022). This is partly due to growing efforts to foster and increase research engagement by consortiums, networks, research projects and libraries in some instances. Thus, collaborations between those who create the data and those who want to use the data is proving to be a proactive solution for increasing scholarly engagement with web archives.

The use of archived web content for scholarly purposes has ongoing pedagogical challenges, due to the characteristics of an archived website or web page which may not be a complete surrogate of what was once on the live web, rather, it is a version (Brügger, 2010). Moreover, the collected web content may undergo technical processes during collection, preservation and to provide access through replay or playback (Brügger 2016, 2018; Schneider et al., 2009). Thus, for Brügger (2019; 2018; 2016) archived web content may be considered as reborn digital media, which is clearly distinct from other types of archived media such as film, television, photographs, and newspapers.

Other commentators note challenges due to the variances between searching on the live web, and searching in a web archive (Costa, 2021; Holzmann & Nejd, 2021; Winters & Prescott, 2019; Jackson et al., 2016b; Nielsen, 2016; Healy et al., 2022). The findings through web archive search techniques also tend to present multiple copies of content captured during different crawls, so they have a temporal dimension, which manifests more challenges. In addition, to preserve a website or web page in its entire capacity to produce meaning, it should be inclusive of links to external (hyperlink) information, and quite often this is not achieved due to selection criteria, acquisition policies, technical glitches, financial constraints,

or legislative and copyright restrictions (Besser, 2000; Schneider et al., 2009; Hockx-Yu, 2014; Milligan, 2019).

Both Brügger (2016) and Schafer (2019) suggest that web archives present challenges due to the “absence” of a traditional style catalogue or registry as an entry point. Costea (2018) identifies a need for improvements to web archives in the areas of discoverability options, data selection, data management, and access to more comprehensive documentation and metadata. Truman (2016) suggests that challenges arise for researchers due to a lack of technical knowledge in the application of data mining techniques to vast volumes of data, as well as a lack of training and experience in using web archives from discovery processes to integrating the use of archived web content with traditional research approaches. Whereas traditional researchers may want to take a more qualitative approach towards using the archived web, they too have challenges due to a lack of research methods and theoretical paradigms for the use of the archived web (Millward, 2015).

Legislation on copyright and legal deposit also presents challenges for scholars and researchers to utilise web archives. Using the UK as an example, scholars discuss the challenges in using UK legal deposit collections which are only accessible on a library terminal in a designated reading room. These challenges include the locked down nature of the library terminal whereby researchers cannot view the source code or copy the URL from the browser, which causes problems for citation, and that users are not allowed to copy and paste text which disrupts the affordances that are used by researchers worldwide (Milligan, 2015; Winters, 2020a). Gooding et al. (2019) also discuss the restrictive nature of UK legal deposit legislation which has no allowance for text or data mining of legal deposit web archive collections, and how this presents a barrier for innovative research. Furthermore, Gooding et al. (2021) suggest that the user was neglected as a stakeholder when it came to drafting the legislation access protocols, which are fundamentally print-centric, and thus, they fail to consider the user in line with digital user expectations, and current trends in information seeking behaviours.

Elsewhere, Maurer (2022) and Healy et al. (2022) note how the provision of onsite ‘only’ access to web archive collections in a designated building makes web archives geographically and socio-economically inaccessible for many researchers. Moreover, Truter (2021) highlights the challenges for web archive end user researchers in terms of access and use of archived web content due to legal restrictions, inclusive of copyright and third-party ownership, privacy policies, and the General Data Protection Regulation (GDPR) in the European Union (EU). This manifests challenges for not only the use of the data, but also affects how and if the data can be made shareable and reusable (Truter, 2021) and runs counter to the

requirement of open science which is being stipulated by a growing number of research institutions and funding agencies (Winters, 2020a). Indeed, Healy et al. (2022) note how “the circumstances (legal, ethical, curatorial, financial, technical, temporal, social, and political) under which an organisation (or individual) archives web collections”, will also affect how end users can access, use, and interpret such collections (Ben-David, 2021; Brügger, 2021; Ogden, Halford & Carr, 2017; Ogden, 2021; Vlassenroot et al, 2019).

Researchers may also be more interested in using big data methods such as topic modelling or network analysis on a web sphere of websites (WARC files) from a specific web archive collection (e.g., Geocities) or to do a longitudinal study across multiple legal deposit annual web domain collections (Milligan, 2019; Brügger et al., 2017; Brügger et al., 2019). However, Maurer (2022) points out that organising large volumes of WARC files for research is difficult for both web archiving initiatives, and end user researchers. Reasons for this are varied and may be “due to a mix of curatorial, technical, legal, economic and organisational constraints” (Brügger, 2021c). This is why Brügger (2021c) stresses the need for solid research infrastructures between the web archives with the data and the research teams wishing to use the data, and this will help overcome some of the legal, ethical, and technical challenges for both communities. Of course, this will require funding, and a cultural shift placing the creator and user as partners in the full web archiving lifecycle.

### Scholarly use of web archives

While the use of web archives for research presents multiple challenges, nevertheless, web archives contain records and documentary evidence of human society and are gradually being recognised and used as resources for the study of the recent past. This is evident in the volume of edited collections and monographs published on the topic in the last number of years alone (Gomes et al., 2021; Brügger & Laursen, 2019; Brügger & Milligan, 2018; Brügger & Schroeder, 2017; Brügger, 2017; Milligan, 2019; Brügger, 2018). In addition, there has been a growing number of journal publications, conference papers, and conference presentations which discuss the use of web archives as resources for research, or which offer case studies in the use of web archives and archived content. These cover topics such as media and journalism, social sciences and ethnographies, public health and telemedicine, information science and law, internet studies, web histories and more.

Literature integrating the use of archived web content for Irish based topics or research is difficult to find, with some exceptions being Malone (n.d.), Harjani (2018), Byrne (2019), Greene & Ryan (2019), Healy (2019), Webster (2019), and Greene (2020). It is also a useful starting point when considering the type of research that has already been undertaken using



web archives for research on Irish based topics, and how it can be built upon. It demonstrates the use of a qualitative approach (Malone, n.d., Healy, 2019), a big data approach (Greene & Ryan, 2019; Greene, 2020) and combining qualitative and big data approaches (Harjani, 2018; Byrne, 2019; Webster, 2019). This provides a good indicator on the types of research which needs to be accounted for in any forthcoming legislation on copyright and legal deposit in the ROI. Moreover, Harjani's (2018) research highlights the important role that social media can play in understanding key events in Irish society, thus, any new legal deposit legislation introduced in the ROI should consider making provisions for the inclusion of social media content.

### Scholarly awareness, engagement, and non-engagement with web archives in Irish academic institutions

Finally, the report offers some insights into the challenges and solutions for using web archives for research in an Irish context through an online survey of lecturers, researchers, and students in Irish academic institutions. The survey sought to investigate the awareness of the existence of web archives for academic research, and engagement and non-engagement with web archives for research purposes in Irish academia. Most prominently, the findings demonstrate a limited awareness of the existence of web archives in Irish academic institutions, and that creating awareness increases the probable likelihood for an increase in researcher engagement. However, the findings suggest that promoting awareness of the existence of web archives by itself may not be sufficient to impact engagement. For an unfamiliar audience, efforts are also needed to demonstrate the importance of archiving the web, the value of web archives for research, and more effort for awareness on how to use web archives for research. The findings also indicate that web archives will become more important for research as time goes on.

The survey findings also present several insights on the challenges that scholars based in Ireland perceive for the future use of web archives and archived web content. How to use web archives and archived web content is presented as a challenge from several outlooks, such as search and navigation, handling large volumes of data, citation practices for using archived web content, and research models for using web archives as a non-established source. Of interest are the different outlooks on the use of large-scale analysis, including the need for training in big data analysis for Humanities researchers in Ireland, but also concerns about the use of big data analysis which does not account for a full understanding of the context of the data. Rather, this might be better achieved with a qualitative approach. This certainly implies that there is a need to consider research models that consider both qualitative and quantitative methods as standalone practices, or a mixture of both as a

combined approach to include web archives as a resource for research in Ireland. The completeness of the data is also mentioned in terms of capture frequencies, as well as challenges with the representativeness of the data in a web archive, in relation to what is presented (or not presented) on the web and what ends up in a web archive. There is also the case that data in a web archive is simply not relevant for a particular research discipline.

On a bright note, the findings show that there is already a small community of web archive users in Irish academic institutions, aged from 18 to 65 years, and at different levels of education and academia. This community utilises web archives and archived web content for coursework purposes, for professional publication and historical research purposes, for teaching purposes, for qualitative and quantitative research purposes, and for access to materials no longer available on the live web. Moreover, because users come from a diverse range of research fields, both multidisciplinary and interdisciplinary deliberation are required to consider the challenges, and potential solutions, for developing research models and paradigms for the use of web archives for Irish based research that are fit for purpose across a broad spectrum of research fields. The user responses also offer some valuable insights on the opportunities for the use of web archives for Irish based research, and there is reason to believe that this community will grow over the next few years, as more academics become aware of web archives as resources for research. However, increases in web archive engagement will also depend on the promotion of awareness of the value of web archives, and demonstrations of use cases in academia as well as the public sphere. Formulating an Irish based multidisciplinary/interdisciplinary research network to comprise of current scholarly users and potential users, web archivists, and information professionals would be of great benefit here. It would assist in addressing potential solutions for developing research models and paradigms for the use of web archives for Irish based research in a broad spectrum of research fields; and enable discussions for the development of frameworks to provide course modules for students in the use of web archives for research, and training courses for educators on how to incorporate web archived content as part of their teaching materials and methods.

To end, the survey findings indicate that awareness of the NLI domain web archive is quite poor, and thus it will warrant a strategy for promotion as a research resource, when it eventually becomes accessible. In this regard it will be essential for the NLI to be afforded the capacity to collaborate with users and develop research infrastructures to help overcome some of the legal, ethical, and technical challenges for both communities (Brügger, 2021c). However, this strategy will require funding, and a cultural shift placing the creator and user as partners in the full web archiving lifecycle. In addition, access to the Irish domain web

archive onsite in the NLI reading room ‘only’ will present geographical and socio-economic barriers for some researchers (Maurer, 2022; Healy et al., 2022). Therefore, in terms of the establishment of an Irish domain web archive, the obvious solution to the access problem would be to make it open access using an ‘Opt-Out’ strategy, however, this is probably unlikely for all types of web content. Therefore, for content that requires restrictions, such as content behind paywalls, there will be a need to consider how access can be provided in more than one geographic location, perhaps in conjunction with other legal deposit libraries across Ireland. Moreover, access provisions should be made for researchers and users who are not affiliated to an academic institution. In the long-term, access should be provided in public libraries across Ireland, and this would ensure that users are not disadvantaged based on geographic location or socio-economic circumstances.

It must also be emphasised that certain categories of websites should be open access by default, including:

- (i) websites belonging to the Irish government, its departments, and its subsidiary agencies, as well as local government and councils,
- (ii) websites belonging to public bodies, quangos, civic agencies, and political parties who receive government funding in any form,
- (iii) websites belonging to owners or organisations who have received funding from the Irish government or any of its subsidiary agencies, and this should be stipulated as part of any funding agreement, and
- (iv) websites which have a variety of Creative Commons licences could also be considered for inclusion for open access.

# 1. INTRODUCTION

Since its invention in the early 1990s, the World Wide Web (the web) has become a major resource for researchers (Day, 2003; Hendler, 2003). Yet it is a transient medium: information is in constant flux with content removal and updates, and the omnipresent '404 Not Found' error. As the early web materialised, concerns about the ephemeral nature of the web also emerged (Brown, 2006; Pennock, 2013). From at least 1994, national libraries and cultural heritage organisations soon realised the need to preserve information and content on the web (Webster, 2017b). In the nascent years of the web, there were increasing difficulties for early search engines to index the vast growth of web content through normal cataloguing techniques (Schneider et al., 2009; Mirtaheri et al., 2013). Subsequently, specially designed software programmes known as web 'crawlers' or 'spiders' started to emerge as a technology to address this from at least 1993 (Mirtaheri et al., 2013), and the development of web crawlers also gave rise to the technology for web archiving (Brown, 2006; Schneider et al., 2009). The underlying reasons for web archiving also stem from existing apprehensions regarding the appraisal, storage and preservation of computational records, electronic information, multimedia and born digital materials in general (Fishbein, 1972; Dollar, 1978; Committee on the Records of Government, 1985; Graham, 1994; Waters & Garrett, 1996; Gardner, 1997; Kuny, 1997).

While web archiving has been around since the mid-1990s, it may be seen as a field that is beginning to mature beyond the establishment phase (Schafer & Winters, 2021; Ben-David, 2021). On the other hand, the use of web archives, and archived web materials for research is much less established, and only saw progress in the past decade or so (Maemura, 2022; Gomes et al., 2021a). Indeed, scholars have highlighted how academics have been slow to embrace web archives as resources for research (Webster, 2020; Rogers, 2019; Leetaru, 2019; Meyer et al., 2017; Webster, 2017b; Winters, 2017; Leetaru, 2017; Brügger, 2016; Meyer et al., 2011; Dougherty et al., 2010). Reasons for this are varied and will be discussed in more detail in [section 2.3](#).

In relation to Ireland, publication of Irish based research integrating the use of archived web content is difficult to find with a few exceptions being Malone (n.d.), Harjani (2018), Byrne (2019), Greene & Ryan (2019), Healy (2019), Webster (2019), and Greene (2020). Also, to date, and to the best of our knowledge, there have been no web archive user studies conducted across Irish academia that examine scholarly engagement, or awareness of the existence of web archives as resources for research. In essence, very little is known about those who engage with, or might potentially engage with, web archives as resources for Irish

based research. Also, to the best of our knowledge there have been no studies which offer an overview of the availability and accessibility of web archives based across the island of Ireland, which would be useful for mapping the opportunities for the use of these resources for research on Irish based topics. This presented an opportunity for further investigation.

The purpose of this report is threefold, (i) to examine the causes for the loss of digital heritage and how this relates to Ireland, (ii) to offer an overview of the landscape of web archives based across Ireland, and their availability, and accessibility as resources for Irish based research, and (iii) to provide some insight into the awareness of, and engagement with, web archives in Irish third-level academic institutions. This report is informed by a collaborative study by Sharon Healy (Maynooth University) and Helena Byrne (British Library) and uses a qualitative exploratory approach, through desk research, a review of resources and relevant literature, and informal dialogues with heritage colleagues. It also engages with a more quantitative approach using an online survey.

To note here, this report refers to Irish digital heritage in the context of the digital heritage of the island of Ireland. When required, we refer to the digital heritage of Northern Ireland or the Republic of Ireland to distinguish between the two jurisdictions.

## 1.1 Background

It is widely agreed that web archiving involves the selection and collection of web content, preserving it for the future and making the collected web content available for access and use (Brown, 2006; Dougherty, 2007; Niu, 2012; Pennock, 2013). While a web archive may also come under the umbrella of a digital archive, it is nonetheless, “a specific type of digital archive” (Lomborg, 2019, p. 99). It is worth noting here that there is a difference between web archiving and website backup. Backup software ensures that an organisation’s website is copied and retrievable in case of data loss or malfunction. It operates at a more present and recent reference point, as earlier backups tend to be overwritten (Bauer, 2018; Crocetti, 2019). Web archiving on the other hand, is a much more complex process (Bingham & Byrne, 2021; Antracoli, 2014; Brügger, 2018) and is representative of the processes and activities described in the Archive-It web archiving lifecycle model which is inclusive of appraisal, selection, capture, storage, quality assurance, preservation and maintenance, replay/playback, access, use and reuse (Bragg & Hanna, 2013; Healy et al., 2022) ([Figure 1.1](#)).

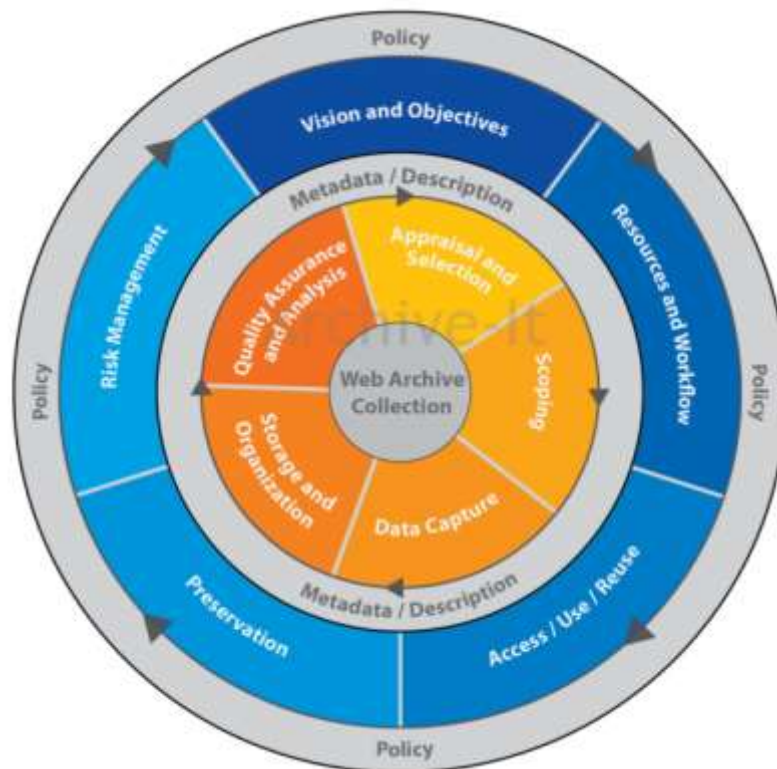


Figure 1.1: Web Archiving Life Cycle Model (Bragg & Hanna, 2013) which is inclusive of appraisal, selection, capture, storage, quality assurance, preservation and maintenance, replay/playback, access, use and reuse

In addition, the terms ‘internet’ and ‘web’ are often used interchangeably, and while they are connected, they are separate entities. In brief, the internet is a networking infrastructure which connects computers, devices, and mobile phones on a global scale, whereas the web is an interlinked information system, operable through the medium of the internet (Beal, 2010; Milligan, 2019, p. 32).

In describing the difference between the internet and the web, Gillies and Cailliau (2000) offer a useful explanation suggesting that the internet is

like a network of electronic roads criss-crossing the planet - the much-hyped information superhighway. The Web is just one of the many services using that network, just as many different kinds of vehicles use the roads (p. 1).

The internet, as we know it today, for the most part operates via Transmission Control Protocol/Internet Protocol (TCP/IP). First, an IP address is an allocated number for a machine connected to the internet, whether it is a laptop, an X-box, or a smart TV. Milligan (2019) suggests thinking of an IP address as a library call number, “letting us quickly locate things in

an otherwise overwhelming sea of items” (p. 36). In explaining TCP/IP further, Computer Hope provides the following explanation.

TCP/IP is a set of rules (protocols) governing communications among all computers on the Internet. More specifically, TCP/IP dictates how information should be packaged (turned into bundles of information called packets), sent, and received, as well as how to get to its destination (Computer Hope, n.d.).

In the Republic of Ireland, the first public connection to the internet (via TCP/IP), went live in Trinity College Dublin in June 1991. University College Dublin also connected several weeks later (Sterne, 2015+). In December 1994, President Mary Robinson became the first Irish Head of State to make use of email “sending Christmas greetings to thousands of emigrants via the Internet” (Cunningham, 1995, p. 17). In February 1995, the Department of Foreign Affairs became the first Irish government department to disseminate an official document via the web, albeit from “a server in the computer applications department at Dublin City University” (Sterne, *IT's Monday*, 1995: Issue 142).

Today, the web is accessed via browsers such as Google Chrome, Microsoft Edge, Firefox, or Safari, and allows for access to web pages which contain a collection of resources and files such as: text, images, graphics, books, newspapers, audio, video, movies, databases, widgets, styling, scripts, software and more (Milligan 2019, p. 4; Day, 2006, pp. 17–18). According to Brügger (2018), the web can be examined through an analytical grid of five strata: an individual web element, an individual web page, an individual website, a web sphere, and the web in its entirety (p. 31). Moreover, this can be applied equally to both layers of the web being “the visible/audible web in the browser, and hidden text of HTML code and associated files” (Brügger, 2018, p. 31). These five strata also offer an equally applicable model for studying the archived web.

For Brügger (2018), there are several types of “active processes” which we might consider to be a form of web archiving as follows:

1. making an image like a screenshot PNG or PDF,
2. making a screen movie or screencast, for example recording a user “moving through a website” or playing a video game,
3. downloading individual files like HTML files, or text extracted from HTML files, or embedded files on a specific web page such as images, audio, or video,
4. web crawling which tends to be used by institutions invested in the large-scale preservation of websites,
5. collecting material through APIs like social media,

6. collecting websites that were taken offline but are preserved intact, for example, a backup of a website stored on a media carrier such as a floppy disk or other types of media carriers,
7. collecting web materials as they appear in other types of media such as books, documentaries, television, or film. Although here Brügger notes that this may not really be classified as there is “no active process taking place” with the content, but worth including as it may be the only source available, especially from the early years of the web (Brügger, 2018, pp. 80–81).

Also coming under the umbrella of web archiving, social media is a challenging online format to archive. The workflows and tools used to capture social media often necessitate a different approach to archiving static, or semi-static pages through web crawling, and may necessitate collecting data through APIs. Collecting social media also present different legal, ethical, and curatorial considerations (Thomson, 2016; Breed, 2019; Bingham et al., 2020; Bingham & Byrne, 2021; Michel et al., 2021; Vlassenroot et al., 2021). Nevertheless, there are several strategies that can be used to collect social media content, in some (but not all) cases.

While Brügger (2018) points to various ways in which the web might be archived (pp. 80–81), and while it is important to acknowledge such efforts, the focus of this report, for the most part, is on institutional web archiving and curation through web crawling, and the use of institutional web archives for research or other purposes.

Web archiving, through ‘crawling’, uses an archival quality web crawler to retrieve and save content from the web through a process called ‘capturing’ or ‘harvesting’. For the most part archival quality crawlers have been developed by the web archiving community themselves. For example, the Heritrix crawler was developed by the Internet Archive from early 2003, and with cooperation from the Nordic based national libraries in the latter part of 2003, Heritrix had its first public release as open source in January 2004 (Mohr et al., 2004, p. 4). Private web archiving companies may also develop their own in-house crawlers. For instance, the company MirrorWeb uses Heritrix, but also developed Electrolyte as a crawler to explore and capture “the most rigorous and dynamic digital domains” (MirrorWeb, n.d.).

Beginning from an initial set of identified Uniform Resource Locators (URLs), known as ‘seed’ and/or ‘target’ URLs, the crawler contacts the hosting web server with a request to retrieve and save the pages that are identified by the URL, the crawler then finds all the hyperlinks on a page which link to other pages, files and assets such as images, PDFs, style sheets, scripts etc., and lists these URLs to the crawl queue. The process is repeated until the queue is emptied, or it reaches a specified URL threshold limit (IIPC, n.d., The WARC Format 1.0;



Brügger, 2018, p. 81; Mirtaheri et al., 2013; Cho & Garcia-Molina, 2000, p. 200). It is also worth mentioning that web curation tends to set the guidelines, rules, and procedures for the selection and collection of web content and ensuring that the web content matches the “curatorial objectives” (Schneider et al., 2009, pp. 210–11). For example, this may involve the development of collection policies, determining the scope for a legal deposit crawl, or making decisions on whether to pursue permissions for external web pages for a selective thematic collection. In some cases, permissions may also need to be sought by an institution which chooses to archive content outside of a national domain for instance, and/or to provide access to content outside of a reading room, as is the case for national libraries in Estonia, New Zealand, and the United Kingdom (UK) (IIPC, n.d., Legal Deposit; Byrne, 2020a).

Capturing content on the ‘surface’ web presents problems due to the wide range of data types such as text, image, sound, visual, multimedia and even software in varied formats “all of which may need to be considered separately from a preservation perspective” (Day, 2006, pp. 17–18). The ‘surface’ web can be accessed by a URL, whereas the ‘deep’ Web is estimated to be far larger, and access often requires encryption keys or an authentication log-in and password, or requires payment for access (Lyman, 2002, p. 41; Day, 2003, pp. 16–17). Bergman (2001) notes how traditional search engines by-pass or cannot retrieve some deep web content as some web pages “do not exist until they are created dynamically as the result of a specific search.” Indeed, to view information on the ‘deep’ web often involves user interaction and the input of a request (Bergman, 2001).

The captured data also needs to be filed and stored. Song and Jaja (2008) suggest that there are a few methods used, but the most popular method for institutional archiving is via “containers in a well-defined structure” such as ARC/WARC formats (p. 2). The Internet Archive ARC file format (ARC\_IA) is a method for merging multiple digital resources together (e.g., HTML, CSS, JPG, PNG etc.) into a self-contained aggregate archival file together with related information (Library of Congress, n.d., ARC\_IA; Burner & Kahle, 1996). Most institutions now prefer the WARC (Web ARChive) file format, which is an extension of the ARC format with revisions to accommodate “related secondary content, such as assigned metadata, abbreviated duplicate detection events, later-date transformations, and segmentation of large resources” (IIPC, n.d., The WARC Format 1.0), although, ARC is still an acceptable legacy format.

In terms of preservation, Day (2006) describes web content preservation as a subset of digital preservation which is concerned with the processes of maintaining captured web content in a usable and accessible condition for the long-term. Web preservation may also be concerned with web archaeology. For Aasman et al. (2019) a “web archaeological” approach focuses on

actively uncovering the history of the web in its early days, emphasising the role of ‘digging’ and ‘reconstructing’ as central methods in tracing material objects (software, hardware, terminals, hard drives, cables, et cetera) and born-digital objects (websites, web elements like banners or avatars, blogs and vlogs, and many other forms of user-generated content) (p. 2)

For some commentators, consideration should also be given to the preservation of software along with the website to preserve the dynamic nature of the website (Alberts et al., 2017; de Haan et al., 2017).

The captured data is then processed to be part of a web archive collection, where access is provided through replay or playback software which offers some form of search interface such as the Wayback Machine, or its open-source counterpart, OpenWayback (Costa, 2021, p. 73). The use and reuse of the captured data completes the sequence of the full web archiving lifecycle. The use of the archived web can be found across various fields of research, covering topics such as internet and web histories, sport history, social, health and political sciences, information sciences, law, media and journalism, religious discourse online, youth and social justice, diasporas, migrant communities, and more (Aasman, 2019; Kahn, 2019; Milligan, 2019; Byrne, 2019; Adelman & Franken, 2020; Gorsky, 2015; Ben-David, 2019; Foot & Schneider, 2006; Cocciolo, 2015; Holzmann et al., 2016; Eltgroth, 2009; Taylor, 2017b; Weber & Napoli, 2018; Bødker & Brügger, 2018; Hofheinz, 2010; Webster, 2019; Webster, 2017a; MacKinnon, 2020; MacKinnon, 2021; Huc-Hepher & Wells, 2021).

Studies which offer examples for the reuse of web archive data are difficult to find, with some exceptions being Ruest (2016), Sherratt and Jackson (2021), Brügger (2021a, 2021b), and Eldakar and Holownia (2022). The reuse of data from web archives is problematic for several reasons, some of which include legal restrictions, inclusive of copyright and third-party ownership, privacy policies, and the General Data Protection Regulation (GDPR) in the European Union (EU) (Truter, 2021).

Finally, the lifecycle of the captured data also includes active maintenance to ensure its long-term digital preservation. The American Library Association describes digital preservation as a combination of

policies, strategies and actions to ensure access to reformatted and born digital content regardless of the challenges of media failure and technological change. The goal of digital preservation is the accurate rendering of authenticated content over time (American Library Association, 2008).

Hence, for many commentators, web archiving is a complex process (Bingham & Byrne, 2021; Antracoli, 2014; Brügger, 2018) which requires a great deal of decision making.

Web archiving decisions must be made on the selection of content to be captured; the technology to use for capturing, storage, preservation, and replay/playback; as well as how to make the collected data accessible for use, and indeed, how flexible this access might be. Furthermore, such decisions may be influenced by social, cultural, and political circumstances; legislations on copyright and legal deposit or lack thereof; and the availability of resources in terms of finance, labour, technology, and organisational infrastructures (Ogden, 2021; Dougherty, 2007; Ben-David, 2019; Ben-David, 2021; Hockx-Yu, 2014; Winters, 2020a; Winters, 2019; Vlassenroot et al., 2019; Brügger, 2021c; Maemura, 2022). For Vlassenroot et al. (2019)

web archiving requires a strategic approach as much is required in terms of technologies, systems, policies, procedures and resources to make web archiving more than merely harvesting and storing online content (p. 86).

As more of the cultural, historical, legal, evidential, informational, and social record happens on the web, heritage institutions are tasked with keeping up with ongoing technological changes to capture and preserve this ephemeral medium (Healy et al., 2022, p. 6). As Truman (2016) points out: “the ever-evolving nature of the web means that the live Web and Internet technology will always be ahead of the capture tools” (p. 20). Because of the enormity of the task, it is at least unreasonable, and probably impossible to expect any one institution to assume full responsibility for archiving everything on the web. Therefore, a multi-agency worldwide approach has materialised (mostly in developed countries) (Gomes et al., 2011) whereby different institutions in different countries endeavour to preserve what they can, and what they deem as relevant for their mandate and stakeholders. In the interests of national heritage, the onus is often on national libraries to save what they can of the national web space inclusive of a selective permissions-based approach, and the routine archiving of the national web domain.

While it may seem inevitable that researchers in the humanities, media studies, and social sciences will integrate archived web content with more traditional formats for research topics from the mid-1990s, nonetheless, scholars have been slow to engage with web archives as resources for research (Webster, 2020; Rogers, 2019; Leetaru, 2019; Meyer et al., 2017; Webster, 2017b; Winters, 2017; Leetaru, 2017; Brügger, 2016; Meyer et al., 2011; Dougherty et al., 2010). For example, Meyer et al. (2011) believed that “the use cases for web archives are not well articulated and have not engaged the research community in any significant way” (p. 4). Dougherty et al. (2010) conclude that there is

a gap between the potential community of researchers who have good reason to engage with creating, using, analysing and sharing web archives, and the actual (generally still small) community of researchers currently doing so (p. 5).

Truman (2016) identifies the need for more communication and collaboration between those who curate, create, and steward web archives, and those who use (or might use) a web archive for purposeful research (p. 3).

In relation to Ireland, publication of Irish based research integrating the use of archived web content is difficult to find with a few exceptions being Malone (n.d.), Harjani (2018), Byrne (2019), Greene & Ryan (2019), Healy (2019), Webster (2019), and Greene (2020). Also, to date, and to the best of our knowledge, there have been no web archive user studies conducted across Irish academia that examine scholarly engagement, or awareness of the existence of web archives as resources for research. Moreover, as pointed out by the web archivist at the National Library of Ireland, “‘It’s difficult to get good analytics on web archive users, due to the fact the selective web archive can be accessed remotely’” (Ryan qtd. in Vlassenroot, 2019, p. 100). In essence, very little is known about those who engage with, or might potentially engage with, web archives as resources for Irish based research. Therefore, it is difficult to assess what types of support and incentives will be most effective for assisting scholars and educators in the use of the archived web for Irish based research and teaching. In addition to this, to the best of our knowledge there have been no studies which offer an overview of the availability and accessibility of web archives based across the island of Ireland as resources for research. First, this would be useful for mapping the opportunities for the use of these resources for research on Irish based topics. Second, it would be useful for assessing support and incentive mechanisms which will foster developments in the use of web archives for Irish based research across a multitude of disciplines, as well as encourage engagement by other types of end users such as public administrators, journalists, legal professionals, web designers, computer scientists, data analysts, and local historians (Healy et al. 2022; Ramesh & Hern, 2013; Winters, 2017; Truman, 2016; Bailey, 2015). This presented an opportunity for further investigation.

## 1.2 Purpose of the Study

As mentioned previously, the purpose of this study is threefold, (i) to examine the causes for the loss of digital heritage and how this relates to Ireland, (ii) to offer an overview of the landscape of web archives based across Ireland, and their availability, and accessibility as resources for Irish based research, and (iii) to provide some insight into the awareness of, and engagement with, web archives in Irish third-level academic institutions.

In pursuit of this, the main objectives for the study are outlined below.

- Examine some of the main causes for the loss of digital heritage and how this relates to Ireland;
- Explore the relationship between legal deposit legislation and the preservation of national heritage, and how this relates to Ireland;
- Raise awareness for the importance of collecting and preserving Irish digital heritage from the internet and the web for current and future generations;
- Document the main challenges for the scholarly use of web archives and how this relates to Ireland;
- Assess the availability and accessibility of web archives based on the island of Ireland that can be used as resources for Irish based research;
- Investigate the levels of awareness of web archives as a resource for study/research in Irish academic institutions, and gain a better understanding of how and why archived web content is used or not used for study/research in Irish academic institutions; and
- Offer some insights which may be useful when it comes to providing support and incentives to assist scholars and educators in the use of the archived web for Irish based research and teaching.

## 1.3 Document Outline

This report is informed by a collaborative study by Sharon Healy (Maynooth University) and Helena Byrne (British Library) and uses a qualitative exploratory approach, through desk research, a review of resources and relevant literature, and informal dialogues with heritage colleagues. It further engages with a quantitative approach using an online survey.

Section 1 offers an overview of the processes and activities of the web archiving lifecycle and discusses the purpose and aims of the study. Section 2 examines some of the main causes for the loss of digital heritage and how this relates to Ireland and explores the relationships between legal deposit legislation and national heritage, and how web archiving has emerged as an activity for the preservation of digital heritage. Thereafter, it focuses on scholarly engagement with web archives, through a review of user case studies and web archive user studies, and discusses the challenges experienced by the web archive user community. Section 3 assesses the accessibility and availability of web archives based on the island of Ireland, and their usefulness as resources for Irish based research with a focus on the PRONI Web Archive, the UK Web Archive and the NLI Web Archive. As part of this, the section synthesises legal deposit history, web history, and political debates to form an understanding of the challenges faced by Irish based web archiving initiatives and their users and examines

how inadequate legal deposit legislation contributes to the loss of Irish digital heritage. Using an online survey, section 4 investigates awareness of web archives, and engagement or non-engagement with web archives in Irish academic institutions and discusses some of the perceived challenges faced by Irish based researchers. In section 5, the report concludes with an overview of the key findings and insights and offers some final thoughts for future work.

## 2. RECOGNISING THE PROBLEMS

*In the past, important parts of our cultural heritage have been lost because they were not archived—in part because past generations did not, or could not, recognize their historic value. This is a cultural problem. In addition, past generations did not address the technical problem of preserving storage media—nitrate film, videotape, vinyl recordings—or the equipment to play them. They did not solve the economic problem of finding a business model to support new media archives, for in times of innovation the focus is on building new markets and better technologies. Finally, they did not solve the legal problem of creating laws and agreements to protect copyrighted material yet at the same time allow for its archival preservation. Each of these problems faces us again today in the case of the Web (Lyman, 2002, pp. 39–40).*

### 2.1 Ephemeral Nature of the Web

In recent years, there has been much recognition for the historical, cultural, informational, intellectual, social, political, journalistic, commercial, and evidential importance of archiving web content (Reyes Ayala, 2013; Brügger, 2018; Milligan, 2019; Dougherty, 2007; Schneider et al., 2009; Weigle, 2018; Foot & Schneider, 2006; Ben-David, 2021; Cowls, 2013; Winters, 2017; Weber & Napoli, 2018; Xie et al., 2013; Denev, et al., 2009; Eltgroth, 2009; Taylor, 2017b). This was not always the case (Masanès, 2006, p. 2). Early on, various schools of thought debated to what extent the web should be archived. Arguments against an excessive approach to web archiving relate to editorial quality issues, in so far as the quality of content on the web was inferior to that which had been appraised through a traditional editing panel (Masanès, 2006, p. 6). For instance, Chakrabarti et al. (1999) noted that a web page might contain “truth, falsehood, wisdom, propaganda or sheer nonsense” (p. 54). Capturing everything also presents issues such as: who has the economic responsibility to collect and preserve the web, to invest in the technology to do so, to provide the storage and preservational maintenance, and to provide the finance for research, development, and training? (Lyman, 2002, p. 39; Grotke, 2011; Taylor, 2011). Because of the enormity of the task, it is at least unreasonable, and probably impossible to expect any one institution to assume full responsibility for archiving everything on the web. Thus, a multi-agency worldwide approach has materialised (mostly in developed countries) (Gomes et al., 2011) whereby different institutions in different countries endeavour to capture and preserve what they can, and what they deem as relevant for their collection mandates and stakeholders. Such institutions include national and regional libraries and archives, university libraries and academic institutions, non-profit organisations, and commercial organisations (Wikipedia, 2011+, List of web archiving initiatives).

Another debate arose within the community of computer scientists who portrayed the web as “a self-preserving medium” (Masanès, 2006, p. 6). Although several studies emerged which would challenge this notion, as the rationale for archiving the web was further reinforced by studies that examine link rot, web content drift, and the extent and frequency of web content change over time. Such studies have been conducted for almost three decades, and span across different disciplines such as education, law, library and information science, information science and technology, computer science, and medical sciences (Harter & Kim, 1996; Koehler, 1999; Lawrence & Giles, 1999; Germain, 2000; Cho & Garcia-Molina, 2000; Lawrence et al., 2001; Markwell & Brooks, 2002; Dellavalle et al., 2003; Fetterly et al., 2003; Hester et al., 2004; Ntoulas et al., 2004; Sellitto, 2005; Goh & Ng, 2007; Wren, 2008; Klein et al., 2014; Zittrain et al., 2014; Zhou et al., 2015; Jackson, 2015a; Bansal & Parmar, 2020; Craigle et al. 2022).

Link rot, also known as reference rot, broken links, or link decay, is used as a term to indicate that a URL no longer provides direct access to a file or web page as originally indicated. Furthermore, even if a URL is stable, the contents of a web page could change; hence, ensuing readers may not view the exact same cited content, or even have the same user experience (Lawrence et al., 2001; Dellavalle et al., 2003; Schneider et al., 2009; Brügger, 2010). For example, computer scientists from Stanford University monitored 720,000 web pages daily over a four-month period and found that 40% of web pages in the .com domain changed their web content daily, while web pages in other domains were at an average of 10% (Cho & Garcia-Molina, 2000, p. 201). Another study by Andy Jackson, examined the URLs of web pages that were archived by the UK Web Archive in 2013 and 2014 to see whether such pages were still available on the live web, or had changed. Jackson suggests that “very few archived resources are still available, unchanged, on the current web. After just two years, 60% have gone or have changed into something unrecognizable” (Jackson, 2015a).

In terms of reference rot, Dellavalle et al. (2003) examined link rot in medical and science journals for URL references and found that, after publication, 3.8% were inactive after three months, which increased to 10% after fifteen months, and 13% at twenty-seven months. Germain (2000) analysed the accessibility of sixty-four URLs cited in thirty-one academic journal articles over a three-year period from 1997-1999, and found 26.5% could not be accessed in 1997, 37.5% in 1998, and 48.4% in 1999. Germain (2000) concludes that for “the scholarly community to retain its integrity, standards must be set to ensure that cited works are retrievable” (p. 364). A study by Spinellis (2003) used two computer science journals to source a sampling of publications from 1995-1999 which cited URL references and extracted 4,375 URL references for verification. Spinellis found that 20% of URLs were inaccessible after



one year of publication, and that this increased from 40% to 50% four years after publication. Spinellis (2003) argues: "Citations in scholarly work are used to build upon existing work [therefore] references that cannot be located seriously undermine the foundations of modern scientific discourse" (p. 71).

For some commentators, the instability of URLs over time may be due to software and system upgrading, changes in filing systems and file names, the re-arrangement of web content, and relocation of servers or server name changes (Berners-Lee, 1998; Besser, 2000; Lyman, 2002; Lawrence et al.; Spinellis, 2003; Pennock, 2013; Masanès, 2006). Other reasons may be due to the relocation of researchers due to an institutional change (Bansal & Palmer, 2020), and the fact that organisations often lose interest in maintaining sites or have not got the time or financial resources to keep sites and URLs up to date (Weisbard, 2011). Tim Berners-Lee, the inventor of the web, suggests that link rot occurs often, and more simply due to a lack of human "forethought" (Berners-Lee, 1998).

Concerns about the ephemeral nature of the web also stemmed from existing apprehensions regarding the appraisal, storage and preservation of computational records, electronic information, multimedia and born digital materials in general (Fishbein, 1972; Dollar, 1978; Committee on the Records of Government, 1985; Graham, 1994; Waters & Garrett, 1996; Gardner, 1997; Kuny, 1997). In the early 1970s, Fishbein speculated that unless archivists were brought up to speed to deal with the challenges of appraising and preserving computational records, "about one million reels of tape in the Federal Government and more elsewhere will be erased without any archival judgments on the continuing value of the information they store" (Fishbein, 1972, p. 35). In 1985, the Committee on the Records of Government (1985) cautioned that the "United States is in danger of losing its memory" due to the shift from paper to electronic records and the instability of maintaining electronic materials (p. 9). The 1996 report, *Preserving Digital Information* by the Commission on Preservation and Access and the Research Libraries Group (RLG) further identified concerns for the preservation of electronic and multimedia materials stressing "the need to protect against both media deterioration and technological obsolescence" (Waters & Garrett, 1996, p. iii, p. 5). For example, media that was stored on nitrate film and magnetic tapes often deteriorated beyond redemption; and information stored on older floppy disk versions were at risk of being unreadable by upgraded technology (Besser, 2000; Lyman, 2002). Terry Kuny (1997) coined the term "Digital Dark Age" to highlight the loss of historical information due to outdated file formats, the upgrading or obsolescence of software and hardware, and the loss of information on the internet.

Reference to the preservation of electronic mail (email) was highlighted by the lawsuit *Armstrong v. Executive of the President* (1989). The lawsuit was first filed to prevent the deletion of emails created by the Reagan and Bush White House administration. Consequently, the case set a precedent for the formal acknowledgement that emails formed a part of the Presidential records to be handed over at the end of term, for appraisal and preservation by the state archivist (Bearman, 1993). Indeed, referring to the Republic of Ireland in 1997, Michael Cunningham of *The Irish Times*, also discusses the preservation of emails, and asks:

If a digital national archive is important for the historians of the future, where is Ireland's digital archive? Which national agency in Ireland should - or could - be responsible for saving and preserving today's email and other electronic objects? (Cunningham, 1997b, p. 18).

While Cunningham notes that “most employees in central and local government” did not appear to have access to or use email technology and points out that the “problem might seem far off”, he suggests that “the longer the State postpones decisions in such areas, the bigger the chunk of our country's digital history that future generations will lose forever” (Cunningham, 1997b, p. 18).

Additionally, in the Republic of Ireland, from 1997 up to 2012, the Director of the National Archives of Ireland (NAI) repeatedly advised on the “pressing need” for the long-term preservation of electronic government records as outlined below.

In the annual reports since 1997, attention has repeatedly been drawn to the pressing need for action to ensure the long-term preservation of records in electronic form. Much of the business of Government is now transacted electronically and it is essential that a legal and regulatory framework, and resources and systems be put in place to ensure that the electronic records generated can be managed and preserved into the future, thereby facilitating Government accountability and preservation of the national memory (NAI, *Report of the Director for 2012*, p. 17).

Moreover, in an article in *The Irish Times* in 2012, the keeper of the NAI, Tom Quinlan highlighted how the Irish Government still did not “have a designated system for preserving and retaining” email (Quinlan quoted in Fagan, 2012).

Regrettably, over twenty years since the problem was identified, the Irish government has still not come to terms with the preservation of electronic records, nor does it seem to have a formal policy for record keeping in any electronic format. This is pointed out, year after year, by the reports of the Director of the NAI from 2014 up until 2020. These reports identify risks

related to the lack of a “comprehensive formal records, management policy for State” and the “Loss of electronic records and archives or access to them, due to degeneration of storage media and/or redundancy of operating systems” (NAI, Reports of the Director for 2014-2020).

On a more positive note, the NAI produced an ambitious strategy in 2021 to deal with the information age, which includes “a digital transformation programme [and] a new framework for records management across government” (NAI, n.d., News; NAI, 2021b). Of course, achieving the goals of the strategy will depend on “improved funding, an enhanced infrastructure and [...] improved staffing resources” (NAI, 2021b, p. 6). It will also depend on the universal adoption of a framework for governmental records management, by civil servants, local government, and the Oireachtas, and this will require an organisational cultural shift which may be more difficult to negotiate (Denning, 2011).

From the mid-1990s, discussions on the preservation of electronic information would further coincide with concerns about the vulnerability of information and documents on the internet and the web. In the spring of 1996, the lead technology officer at Microsoft, Nathan Myhrvold, started an email conversation to bring attention to the loss of historical information and evidence, due to the disappearance and replacement of websites and documents on the web, and the turnover or deletion of bulletin board system (BBS) newsgroups on the internet (Gardner, 1997, p. 3). In 1997, Michael Cunningham of *The Irish Times* described the early web as being somewhat,

likened to one vast, rapidly fluctuating library (of bits rather than atoms). But unlike a traditional library it is being rebuilt every minute. Its sites can flicker and die in days, hours or even seconds. Web pages are revised and spiced up with fancier graphics and revamped designs, more "plug-in" animations and "applets", often with no record kept of the previous mutations (Cunningham, 1997b, p. 18).

Other examples relate to government websites. In 2003, David Worlock of Electronic Publishing Services Ltd. in London claimed that 25% of the 2,483 British government websites change their URL every year (qtd. in Weiss, 2003). At the time, Worlock contended that it was problematic as some government documents only existed as a web page, an example being that the dossier produced by the British government on Iraqi weapons only ever appeared as a web page. Thus, Worlock suggests that there is “no definitive reference where future historians might find it” (qtd. in Weiss, 2003). More recently, Brügger (2018) remarks that at the time of the inauguration of Trump, there were substantial changes to the official White House website ([www.whitehouse.gov](http://www.whitehouse.gov)), including the removal of topics on climate change and global warming which had been published on the site by the former President, Barack Obama (p. 1). Paul Koerbin of the National Library of Australia also refers to the transient nature of

websites during election campaigns, changes of government, party leadership challenges and government leadership changes (Koerbin, 2013a; Koerbin, 2013b). In the Republic of Ireland each change of government leads to changes in departmental titles and often major reallocation of ministerial responsibilities, and up until recently, changes in departmental titles usually entail the creation of new URLs for departmental websites (Healy, 2016; see [Figure 3.12](#)).

The main Government of Ireland website ([www.irlgov.ie](http://www.irlgov.ie)) also saw several redesigns from 1997-2008, and a new URL ([www.gov.ie](http://www.gov.ie)) sometime in late 2008 with, of course, a new redesigned website, followed by more redesigns and broken links. Michael Cunningham of *The Irish Times* offers an overview of the main Government of Ireland website after it was redesigned in March 1997 as follows:

THE Government's Web site has just undergone a spring clean after almost a year online. Still at that same awful Web address, now the interface is 'frames-biased' - which means most users can't download or bookmark individual pages apart from the main menu. The revamped design is a post modern dog's dinner, mixing last year's navigational icons with a new set, and a mishmash of backgrounds - some are the old harvest yellow, others (for no apparent reason) are dull default greys. And the frustrating feature of extra, redundant pages between the link you click and the - 'final' Web page you want is still there [...]

[There] still are no daily Dail/Senate debates, and the policy about digital versions of other documents is very inconsistent [...] And when Ministers mention that documents are online or tell us their email addresses, often the information isn't a 'hot link' [...]

The search engine is still bizarre too, suffering from a bout of election madness. Search for documents with the word 'abortion' in them, and you get 'Provisions Relating To Petroleum Taxation'! (Cunningham, 1997a, p. 10).

[Figure 2.1](#) offers an overview of some screenshots of the archived web pages in the Wayback Machine showing the changing nature of the Government of Ireland website from 1996 to 2011.



Timestamp: 1996-12-24  
(www.irlgov.ie)



Timestamp: 1997-01-05  
(www.irlgov.ie)



Timestamp: 2000-03-02  
(www.irlgov.ie)



Timestamp: 2002-03-28  
(www.irlgov.ie)



Timestamp: 2008-12-07  
(www.gov.ie)



Timestamp: 2011-07-03  
(www.gov.ie)

Figure 2.1: The changing nature of the Government of Ireland website captured in the Wayback Machine from 1996 to 2008 (www.irlgov.ie), and 2008 to 2011 (www.gov.ie)

In December 2017, the Irish government decided to migrate all the departmental websites under one main government website. Starting with the website of the Department of An Taoiseach, departmental websites began migrating to the new centralised website in 2019. The Taoiseach at the time, Leo Varadkar (Fine Gael) outlined the purpose of the plan in Dáil Éireann, as follows.

Departments are currently represented online by multiple distinct websites and platforms, each providing different visual styles and user experience. A Government decision was taken in December 2017 to migrate all primary Department websites to one single portal, gov.ie. This aligns with international best practice. Gov.ie has been developed with the citizen at its centre, with an emphasis on policy and service areas, as opposed to how a Department is structured internally. [...]

My Department was the first to migrate to gov.ie, with more to follow on a quarterly basis over the next 18 months. Our new web address is www.gov.ie/taoiseach. The next to migrate across will be the Departments Public Expenditure and Reform, Finance and Rural and Community Development, and they are now in the process of transitioning. All remaining Departments will transition between now and the end of next year [...]

Users of the *taoiseach.ie* website were notified in mid-November 2018 that the information on the *taoiseach.ie* website would move to the new central *gov.ie* website. The old *taoiseach.ie* website was available until 22 February and has since been archived in co-operation with the National Library of Ireland (Leo Varadkar, Dáil Éireann, Departmental Websites, 26 February 2019, italics in original).

The decision to centralise government department websites into one main website, caused some debate on why the government had not done more market research into the decision and surveyed the users to see what they wanted (Micheál Martin, Dáil Éireann, Departmental Websites, 26 February 2019). Nonetheless, it is reassuring to see that the government liaised with the National Library of Ireland to archive the websites before they started the migration. One would also hope that the government will also liaise with the library regarding the “archivability” of their new website. Stanford Libraries describe archivability as “the ease with which the content, structure, functionality, and front-end presentation(s) of a website can be preserved and later re-presented, using contemporary web archiving tools” (Stanford Libraries, n.d., Archivability).

In terms of national responsibilities for the collection and preservation of national heritage, it is worth noting here that national archives tend to collect and preserve the ‘records’ and documents of government departments and agencies, and may include the collection of records and documents from public bodies, ombudsman agencies or quangos, etc. National libraries tend to collect and preserve the ‘publication’ outputs of a nation state, through a system called legal deposit, which is usually a statutory obligation on publishers to deliver a copy of all new publications. However, libraries may also have special archival collections containing ‘records’ and documents, while archives may have libraries containing specialised bibliographic collections or rare books. It is widely acknowledged that the collection efforts of national archives and national libraries constitute a major contribution to the preservation of national heritage (Yusuf, 2013; James, 2019; Larivière, 2003; Gooding et al., 2019; Arnold-Stratford & Ovenden, 2020).

According to UNESCO (2003) “the disappearance of heritage in whatever form constitutes an impoverishment of the heritage of all nations”, and digital heritage should not be an exception. In the Charter on the Preservation of the Digital Heritage, UNESCO (2003) characterises digital heritage as “unique resources of human knowledge and expression” which embrace

cultural, educational, scientific and administrative resources, as well as technical, legal, medical and other kinds of information created digitally, or converted into

digital form from existing analogue resources. Where resources are ‘born digital’, there is no other format but the digital object (UNESCO, 2003).<sup>1</sup>

Some of the factors which contribute to the loss of digital heritage to posterity include technological obsolescence of hardware and software, availability of resources, and “the lack of supportive legislation” (UNESCO, 2003). There is also the case that “Attitudinal change has fallen behind technological change” and consequently, the “threat to the economic, social, intellectual and cultural potential of the heritage - the building blocks of the future - has not been fully grasped” (UNESCO, 2003).

While web archiving offers a solution for nation states to capture and preserve their digital heritage on the web, it is often the case that governments and societies have not quite grasped the reality of the consequences for the loss of this heritage and therefore lack the urgency to develop strategies for its collection and preservation. The focus is on the creation of new information and not the preservation of the old (Lyman, 2002, p. 39). Indeed, Morris (2019) suggests that the loss of digital heritage is often due to realising “too late that the latest technology is neither mundane nor permanent” (Morris, 2019, p. 500). But, even if the concerns for the preservation of digital heritage on the web is acknowledged, it might take “decades until the technical, organisational, and economic conditions are in place to launch preservation initiatives” (Brügger, 2019, p. 16). We might also add political and legal “conditions” to this list, as they present further barriers for the launch of preservation initiatives and will be discussed in more detail in [section 2.3](#), and [chapter 3](#). However, to note here, the challenges with the long-term preservation of digital heritage are not unique to the internet or the web.

Lyman (2002) notes how societies have lost important parts of their cultural heritage in the past because it was not archived or preserved due to “cultural”, “technical”, “economic”, and “legal” problems (pp. 38–39). Lyman (2002) points out that this is

in part because past generations did not, or could not, recognize their historic value. This is a cultural problem. In addition, past generations did not address the technical problem of preserving storage media—nitrate film, videotape, vinyl recordings—or the equipment to play them. They did not solve the economic problem of finding a business model to support new media archives, for in times

---

<sup>1</sup> With this in mind, the National Library of the Netherlands are responsible for collecting and curating websites hosted by the former Dutch internet provider XS4ALL, to become the “the first born digital collection in the world” to receive a place on UNESCO’s register for Documentary Memory of the World (Teszelszky, 2022).

of innovation the focus is on building new markets and better technologies. Finally, they did not solve the legal problem of creating laws and agreements to protect copyrighted material yet at the same time allow for its archival preservation. Each of these problems faces us again today in the case of the Web (pp. 38–39).

While Lyman may have published this summary in 2002, it is still applicable today, although one could argue that the web archiving community has come a long way in providing solutions to the “technical” problem. On the other hand, web archiving is complicated by “ever-evolving” internet, web, and software technologies, thus, such technologies “will always be ahead of the capture tools” (Truman, 2016, p. 20). Nonetheless, continual efforts are being made by heritage organisations and web archive curators to capture what they can, as best they can (Laursen & Møldrup-Dalum, 2017, p. 220).

From at least 1994, libraries, archives and cultural heritage organisations have also had concerns about the ephemerality of web content. The National Library of Canada (now part of Library and Archives Canada) initiated discussions in 1994 around the collection of electronic materials, inclusive of websites; and initiated a pilot project in 1995 (Webster, 2017b, p. 177). The National Library of Australia organised a working group to address collection and archiving techniques for the web in 1995 and initiated a web archiving programme in 1996 (Schneider et al., 2009, p. 206). At the same time, the development of web crawler programmes gave rise to the technology for web archiving (Schneider et al., 2009, p. 206). Research by Gomes et al. (2011) provides an overview of global development in web archiving initiatives. This work also provides the base for a Wikipedia article. The article is regularly updated to document the growing number of global web archiving initiatives. It shows a significant increase in web archiving activities by non-profit organisations, commercial organisations, academic institutions and national and regional heritage organisations and associations, with much of the growth occurring in North America and Europe.

Some of the earliest efforts in establishing web archiving initiatives may be attributable to the National Library of Canada from 1995, the Internet Archive from 1996, the National Library of Australia (PANDORA) from 1996, the Smithsonian Museum in 1996, and the Royal Library of Sweden from 1997 (Webster, 2017b; Brown, 2006; Koerbin, 2021; Milligan, 2019; Arvidson et al., 2000). Several national libraries followed suit to develop web archiving programmes such as the National Library of New Zealand (1999), the National Library of the Czech Republic (2000), the Library of Congress (2000), the National Library of Korea (2001), the National and University Library of Croatia (2004), and the National and University Library of Iceland (2004) (Wikipedia, 2011+, List of Web Archiving initiatives).



It is also worth mentioning the work of the International Internet Preservation Consortium (IIPC). The IIPC was founded in 2003 by twelve members, being the national libraries of Australia, Canada, Denmark, Finland, France, Iceland, Italy, Norway, Sweden, The British Library (UK), The Library of Congress (USA) and the Internet Archive (USA). The IIPC set out to achieve several goals as outlined below, which are extracted ‘verbatim’ from an archived copy of the IIPC about/index page from 2004 in the Wayback Machine (Figure 2.2):

- To enable the collection of a rich body of Internet content from around the world to be preserved in a way that it can be archived, secured and accessed over time.
- To foster the development and use of common tools, techniques and standards that enable the creation of international archives.
- To encourage and support national libraries everywhere to address Internet archiving and preservation.



Figure 2.2: Screenshot of the IIPC about/index page, 2004, captured in the Wayback Machine (Timestamp: 2004-06-03 01:41:15)<sup>2</sup>

Today, the IIPC has over fifty members from thirty-five countries, and has broadened its goals over the years to further include:

- developing “international advocacy for initiatives and legislation”,
- fostering “broad international coverage in web archive content through outreach and building curated collaborative collections”, and

---

<sup>2</sup> IIPC about/index page, Wayback Machine, 2004,  
<https://web.archive.org/web/20040603014115/http://netpreserve.org/about/index.php>

- encouraging and facilitating “research use of archived Internet content” (IIPC, n.d., About Us).

The IIPC curated collaborative collections are hosted on the Archive-It platform, and cover topics such as the Summer and Winter Olympics, the Summer and Winter Paralympics, the European Refugee Crisis, and COVID-19 (IIPC, Archive-It, <https://archive-it.org/home/IIPC>). Also, the IIPC and their members responded to other global events to capture and preserve web content during the Ukrainian war (IIPC, 2022, IIPC webinar). This is a good example of how political conditions might influence the selective processes for a web archive collection. The IIPC organises an annual web archiving conference which is one of central events in the calendar of the web archiving community and has seen a dramatic rise in the contribution by individuals who use the archived web for research and is evident from their programme schedule from May 2022 (IIPC WAC 2022 program, <https://netpreserve.org/ga2022/wac/>).

## 2.2 From Print to Non-Print

Due to developments in computer technology and information communications, the concept of electronic publishing was realised. Electronic publishing also brought with it a new dimension for legal deposit schemes, which up to this point had been mostly print-centric (Muir, 2005, p. 4). In the 1960s, electronic publications referred to the use of computers to produce print publications using word processing, typesetting, or mark-up tools. In the 1970s, the first example of an electronic journal was distributed as “a computer readable archival file” and “in the form of computer-output microfiche” (Lancaster, 1995, p. 520). Indeed, microfiche grew as a media format for publication outputs (Schafer, 2020). The development of the CD-ROM in the 1980s offered an effective, low-cost solution for e-publishing and allowed for good quality graphics and images (Pettenati, 2001). From the early 1990s, internet journals emerged, and email technology allowed for the distribution of e-publications via mailing lists, though this was originally in plain text format (Lancaster, 1995; Pettenati, 2002). CD-ROM began to be replaced as a medium from the end of the 1990s, due to the development of the web, and the increasing availability of the internet (Waniata, 2018).

Today, an electronic or digital publication can refer to a text encoded version of a print publication (e.g., encoded using XML/TEI), a scanned/digitised version of a print publication (e.g., a PDF or microfiche), or a born digital publication where there is no print parallel. Examples here include web pages and blogs on the web; e-zines and e-newsletters distributed via email technologies; information disseminated on bulletin board systems (BBS) via the internet; social media delivered through web or mobile phone technologies; and even interactive CD-ROMs (e.g., CD-ROM encyclopaedias) ([Figure 2.3](#)). It also includes other types

of publications that can be hosted on the web such as podcasts, videos, digital scholarly editions, interactive databases containing bibliographies, statistics, spatial data etc. (Boston, 1998; Taylor, 2013). One must also consider here that each of the publication types mentioned above come with their own form of “digitality” (Brügger, 2019, p. 17).



Figure 2.3: Screenshot of the homepage for the web version of the interactive CD-ROM ‘Safeguarding our documentary heritage’ created by UNESCO and IFLA-PAC which provides interactive guidelines on the preservation of archival materials including multimedia formats, captured in the Wayback Machine (Timestamp: 2001-04-29 03:33:55).<sup>3</sup>

Brügger (2019; 2018; 2016) examines how various types of digital media (materials) come with their own form of digitality, “that is, a specific way of being digital” (2018, p. 5). In this respect, Brügger (2019; 2018; 2016) puts forward three major categories for distinguishing different types of digital media. Brügger (2019; 2018; 2016) offers the typologies of: digitised media, as media that is originally non digital (such as analogue materials that have been digitised, e.g., paper or parchment documents, print newspapers, print photographs, negative films, or any “form of transformation of analogue material into digital form”

<sup>3</sup> UNESCO and IFLA-PAC, Safeguarding our documentary heritage, Wayback Machine, 2001, <https://web.archive.org/web/20010429033355/http://webworld.unesco.org/safeguarding/en/index.html>

(Brügger, 2016, para 28). Then there is born digital media, which is media that has only ever existed in a digital form (such as material on a CD, DVD, the internet, or the web); and reborn digital media, as media that has been collected and preserved and has undergone a change due to this process, such as emulations of computer games or materials in a web archive (Brügger, 2019; 2018; 2016).

As a result of emerging publishing technologies several countries began to review and amend their copyright and legal deposit legislations to incorporate the deposit of non-print materials such as electronic publications stored on devices like CD-ROMs or published online, as well as the archiving of national web domains at scale. For example, in Denmark, legal deposit for print publications has existed since 1697, through a Royal Ordinance, which mandated publishers to deposit five copies of everything they printed in the Royal Library (Dupont, 1999, p. 244). Thereafter, the law was updated on several occasions for the inclusion of maps, and an extension of legal deposit to include Danish territories, for example. A new law passed in 1997 brought with it “a revolution in the history of legal deposit” as it requested “not only printed material, but all works published in Denmark, regardless of the medium used for the production of copies” including “published works on the Internet, that form a final and independent unit and which have been produced for a Danish audience” (Dupont, 1999, p. 245). The legislation was further revised in 2004 to broaden the scope for the inclusion of archiving the Danish national web domain (Webster, 2017b, pp. 179–180). Web archiving is therefore a legal obligation on the part of some, but not all legal deposit institutions. For example, in the Republic of Ireland, digital legal deposit was enacted through the Copyright and Other Intellectual Property Law Provisions Act 2019, which allows for the collection of e-books and online journals by the National Library of Ireland and other nominated legal deposit libraries. However, the legislation does not allow for the routine archiving of the Irish national web domain (Ryan et al., 2022). The collection of the web space of Northern Ireland is covered by UK legal deposit through The Legal Deposit Libraries (Non-Print Works) Regulations 2013 which enables the routine archiving of the UK national web domain. This will be discussed in more detail in the next chapter.

While the histories of copyright and legal deposit are often intertwined, they have had different trajectories and interpretations. For example, the concept of copyright only evolved following the invention of the printing press, and publishers-printers were the first to feel the need to protect their rights to publish a specific work by an author, and prohibit other printers replicating the work, to sell at a lower price (Matthews, 1890, pp. 587–588). The ‘concept’ of a legal deposit scheme for print publications was first implemented in France in 1537 through the *Ordonnance de Montpellier*, a royal decree by King François I of France. Partridge (1938)

notes that the idea of collecting a rich body of literature, at no cost to the state soon spread to other monarchies (p. 3).

In Great Britain, the first implementation of the concept was introduced in 1610 by Sir Thomas Bodley, the founder of the Bodleian Library at Oxford University in 1602. Bodley negotiated a deal in 1610 with the Company of Stationers (London), for them to supply a copy of every new book to the Bodleian Library, published by its members (Muir, 2005; Larivière, 2000; Feather, 1994). Similar efforts in the concept of deposit schemes for print publications were introduced in Sweden (1661), Poland (1645), Denmark (1697) and Finland (1702) (Larivière, 2000; Muir, 2005). The development of more elaborate legislations regarding the copyright of print publications, and by extension legal deposit, evolved over the centuries on a national basis.

In general, most legal deposit schemes require producers of print publications to deposit a copy of each new publication in a nominated institution, often designated as a national library, state library or university library (Gooding et al., 2019). According to Healy et al. (2022) legal deposit schemes serve as a system “to compile, maintain, and provide access to a comprehensive collection and bibliographic record of a country’s published output and, in doing so, creates a significant manifestation of national cultural heritage” (p. 4).

For Lariviere (2000),

Legal deposit legislation serves a clear national public policy interest by ensuring the acquisition, the recording, the preservation and the availability of a nation’s published heritage. Such a national collection is undoubtedly one of the major components of a country’s cultural policy and should also be considered as the foundation of a national policy of freedom of expression and access to information (p. 4).

Today, many countries have legal deposit legislation which enables the archiving of non-print publications such as e-books, audio books, and e-journals, as well as the archiving of the country’s national web domain. This might include the deposit of such publications via physical carriers such as a CD-ROM. Archiving a national domain tends to include a crawl of a country code top level domain (ccTLD), without having to request permissions from website owners. However, defining the scope of a national domain is problematic. For example, if a domain web archive is based solely on a ccTLD like .ie or .uk, it will inevitably exclude thousands of websites with domains such as .com, .org, or .net (Day, 2003). Day (2003) further suggests that national domain archives might also include websites from servers that are physically located in a country, or websites that belong to organisations or individuals when “the intellectual content is of relevance” for national digital heritage (p. 16).

Access to national domain legal deposit collections is also dependent on the legislation, and for the most part, may only be accessible onsite in a designated building or reading room. There are some exceptions such as the Croatian and Icelandic national domain web archives, which are both available online as open access (Winters, 2020a). On the other hand, while the Swedish legislation allows for the collection of the Swedish domain, there is currently no provision for access, and the Danish web archive can only be accessed offsite by researchers with support from an academic institution on a project permissions basis (IIPC, n.d., Legal Deposit; Winters, 2020a). Countries that do not have the necessary legal deposit legislation for archiving national domains, have few options but to engage in collection development based on fair use, or through permissions-based selective collection only, which is a resource intensive process (Costa, 2021).

On top of this, there are other challenges with permissions-based selective collections. Not all websites provide contact details and even if a contact is found there is no guarantee that a website owner will respond (Ryan et al., 2022; Healy et al., 2022; Byrne, 2020a). Byrne (2020a) highlights the challenges of identifying contact details to request open access permission for content archived by the UK Web Archive and suggests that even when contact details are identified and permission requests are sent to content owners, there is a very high failure rate. On average only 20% of requests sent by the UK Web Archive have resulted in open access permission being granted. However, this is a general figure across the whole archive, the actual figure for individual curated collections varies depending on the type of content that is selected. As of June 4, 2018, the response rate for the 'Sport: Football' collection was at 10.49% (Byrne, 2020a, p. 5).

Pennock (2013) further discusses a weakness of selective web archiving due to "the possible or unintentional and unacknowledged selector bias" (p. 9). Pennock (2013) explains that the selection of websites is:

commonly a manual process that reflects the particular interests or knowledge of the person(s) choosing sites for the collection. The sheer size of the Internet, the number of websites hosted and the speed at which information can be published, all make it very difficult for manual selectors to keep abreast of new sources, especially for event-based collections (p. 9).

Brown (2006) further points out that "the greater the degree of selectivity employed, the more subjective the resultant collection will be, constraining the as-yet-unknown requirements of future researchers" (p. 32). Thus, legal deposit libraries tend to conduct both selective and domain-wide web archiving as a more balanced and representative approach towards the capture of national digital heritage.

Other global web archiving initiatives archive the web on a fair use basis, regardless of borders, such as the Internet Archive, a non-profit heritage institution based in San Francisco. Rated as one of the largest web archives in the world, the Internet Archive began web archiving in 1996, and currently provides “unrestricted access” to their web archive through the Wayback Machine (Webster, 2017b, pp. 176–177; Brown, 2006, p. 9). The Wayback Machine, as a search interface, was not publicly available until 2001 (Rogers, 2013, p. 65). The Wayback Machine is globally accessible online and allows users to save a web page by inputting a URL to be captured in real-time. While the Internet Archive collects on the grounds of fair use, it provides an opt-out, takedown clause for website owners, who do not want their websites available in the public Wayback Machine (Lowcock, 2020). The clause states: “The Internet Archive may, in appropriate circumstances and at its discretion, remove certain content or disable access to content that appears to infringe the copyright or other intellectual property rights of others.” Website owners then need to provide relevant details to the Internet Archive inclusive of a statement “made under penalty of perjury” that the information they provide “is accurate” and that they “are the owner of the copyright interest involved or are authorized to act on behalf of that owner” (Internet Archive Help Center, n.d., Wayback Machine General Info).

## 2.3 Web Archives and Scholarly Engagement

As mentioned previously, while engagement with web archives for scholarly research purposes has developed in the past decade or so (Maemura, 2022), nevertheless, several commentators observe how scholars were slow to embrace web archives as a research resource (Webster, 2020; Rogers, 2019; Leetaru, 2019; Webster, 2017b; Winters, 2017; Leetaru, 2017; Meyer et al., 2011; Dougherty et al., 2010). In a bid to gain more understanding of the challenges and opportunities for scholarly engagement with web archives, the report reviews the literature on using the archived web for research and examines a selection of web archive user studies.

### 2.3.1 Challenges for scholarly engagement with web archives

In terms of the slow engagement with web archives by researchers, there are several reasons put forward for this with the most obvious being a lack of awareness of the existence of web archives, or simply because some academic disciplines have no need to rely on such sources (Jatowt, 2008; Riley & Crookston, 2015; Winters, 2017; Costea, 2018; Healy, 2019; also see [section 4.4.5](#)). Initially, web archiving initiatives tended not to prioritise how web archives would or might be used (Dougherty et al., 2010; Hockx-Yu, 2014; Schroeder & Brügger, 2017;

Gooding et al., 2021). Rather, Thompson (2008) suggests web archiving institutions “followed a traditional model of acquisition where material is held in the belief that it has value even though there may be no immediately identified user” (pp. 19–20). For example, the New Zealand National Library did not examine the extent of awareness and use of their web archive before late 2014, even though they commenced a web archiving initiative in 1999 (Riley & Crookston, 2015). Other surveys of the procedures, practices and policies of web archiving institutions show similar tendencies. The National Digital Stewardship Alliance, Content Working Group (2012), found “an area of uncertainty” by web archiving institutions vis-à-vis how collections were being used (p. 11). Schroeder and Brügger (2017) also note that for many years, web archiving initiatives struggled

to set up archiving procedures, hardware and software to keep pace with the seemingly endless flow of new web content and ever evolving software development, while little attention was paid to who might use the material in the archive, and how it might be used (p. 12)

In a Harvard Library report, Truman (2016) stresses the need “for greater communication and collaboration” with researcher communities, as well as the “the need to gather researcher feedback on requirements and impediments to the use of web archives” (p. 42). Although, to do this, one might need to identify a community of users, or even potential users in order to attain user feedback (see for example, Ras & van Bussel, 2007; Stirling et al., 2012; Gooding et al., 2019). Thus, Truman (2016) identifies the need for more communication and collaboration between those who curate, create and steward web archives, and those who use (or might use) a web archive for purposeful research (p. 3). Certainly, it could be argued that a lack of dialogue or collaboration between the creators of web archives, and end users (or even potential end users) has had some effect on engagement with web archives for research purposes. This needs to be addressed going forward.

On a positive note, collaboration between web archive creators and end user researchers has been improving over the past decade (Schroeder & Brügger, 2017; Webster 2017b; Maemura, 2022). This is partly due to growing efforts to foster and increase research engagement by consortiums, networks, research projects and libraries in some instances. Examples include: the International Internet Preservation Consortium (IIPC), the Research Infrastructure for the Study of Archived Web Materials (RESAW), the Analytical Access to the Domain Dark Archive (AADDA) project, the Big UK Domain Data for the Arts and Humanities project (BUDDAH), the Web90 project (Web90: Heritage, Memory and History of the Web of the 1990s), the Web Science and Digital Libraries Research Group at Old Dominion University (ODU WS-DL), the Web ARChive studies network researching web domains and events (WARCnet), ResPaDon



(network to develop and diversify the uses of web archives) and the Archives Unleashed project. In addition, the Archives Unleashed project launched a programme to provide support for cohorts, to further foster research engagement (The Archives Unleashed Project, n.d., Archives Unleashed Cohorts), while Arquivo.pt developed an annual award for initiatives which use and demonstrate the use of the Arquivo.pt web archive (Arquivo.pt, n.d., Arquivo.pt Awards).

The conference programmes of organisations like the IIPC and RESAW often feature workshops for end users and researchers who engage with web archives. The UK Web Archive also provides support for researchers and PhD students using its collections and has been proactive in collaborating on workshops and training. Indeed, the IIPC, libraries and other like-minded organisations often collaborate to provide seminars and workshops for web archive users/researchers, some of which include working with big digital data. Although, when it comes to handling such data, there is often a prerequisite to have some programming skills. For example, in a call for participation of researchers at an Archives Unleashed 4.0: Web Archive Datathon held in the British Library in June 2017, it was suggested that: “Researchers should be comfortable with command line interactions and knowledge of a scripting language (such as, but not limited to Python) is strongly desired” (IIPC members list, Email, 04 April 2017). While this is certainly a good thing for those who are comfortable with programming, it might also be seen as a barrier for entry by scholars with limited technical skills (Bingham & Byrne, 2021).

There is also the need to consider that some academics are more comfortable with, and trusting of, ‘proven’ traditional research methods, although this is not something unique to web archive research. Other disciplinary fields that have had some history of engagement in the use of computational methods for big data analysis, have had similar experience with traditional researchers being mistrustful of computing methods. Examples include humanities computing, history and computing, and the development of computational methods in the social sciences (Drucker, 2012; Hindley, 2013; Winters, 2018; Kelle, 1997). Indeed, in the social sciences, it was only with the rise of the personal computer in the 1980s that computational methods for qualitative research began to gain any kind of traction in the academy, despite the availability of software and tools since the late 1960s (Kelle, 1997).

Other challenges arise, due to the characteristics of an archived website or web page which may not be a complete surrogate of what was once on the live web, rather, it is a version

(Brügger, 2010, pp. 6–7).<sup>4</sup> Deficiencies in the archived artefacts may occur because of the temporal dimensions such as the time it takes to capture, and the possibility of content updates during capture. Deficiencies may also occur due to technical issues such as glitches during the archiving process such as robots.txt or limitations with the archiving software/hardware to keep up with the constant change and upgrade of web media file types and the evolving nature of dynamic content (Brügger, 2010; Meyer et al., 2011; Pennock, 2013; Maemura, 2018; Bingham & Byrne, 2021).<sup>5</sup> For example, Morris (2019) and Aasman (2019) discuss the absences of sound and audiovisual content in web archives. Aasman (2019) notes how the Wayback Machine is “unable to reproduce flash-based videos” and thus, early captures of YouTube pages “show nothing but a front page with empty screen” (p. 43). Morris (2019) also draws attention to the challenges for finding sound files in a web archive. First, while there may be an icon displayed for a sound file on an archived webpage, the actual audio file may not have been captured (p. 497). Second, “preserving audio formats often require preserving the sounds themselves as well as the technologies on which to play those sounds.” Thus, even if the file is there, there may be no way to open it or play it without the obsolete software which created it (Morris, 2019, pp. 497–499).

In addition to the challenges above, in order to preserve a website or web page in its entire capacity to produce meaning, it should be inclusive of links to external (hyperlink) information. Quite often this is not achieved due to selection criteria, acquisition policies, technical glitches, financial constraints, or legislative and copyright restrictions (Besser, 2000; Milligan, 2019; Hockx-Yu, 2014). Furthermore, if hyperlinks direct to social media sites, this presents an additional set of technical, ethical, and legal challenges (Breed, 2019; Bingham et al., 2020; Bingham & Byrne, 2021; Vlassenroot et al., 2021). Finally, the collected web content may undergo technical processes during collection, preservation and to provide access through replay or playback (Brügger 2016, 2018; Schneider et al., 2009). Thus, for Brügger (2019; 2018; 2016) archived web content may be considered as reborn digital media, which is clearly distinct from other types of archived media such as film, television, photographs, and newspapers. Consequently, Brügger (2018) suggests that “historians have to become familiar with this type of source, its characteristics, and how these characteristics impact its

---

<sup>4</sup> Brügger and Finnemann (2013) propose that the archived web is “a Reborn, Unique and Deficient Version and Not Simply a Copy of What was Once Online” (p. 74).

<sup>5</sup> Robots.txt, also known as the robots exclusion standard, or robots exclusion protocol, refers to a standard that is used in websites “to indicate to visiting web crawlers and other web robots which portions of the website they are allowed to visit” (Wikipedia, 2002+, Robots exclusion standard).

scholarly use”(p. 3). Therefore, this implies that the use of archived web content for scholarly purposes has ongoing pedagogical challenges.

Other commentators describe challenges with web archives due to the differences between searching on the live web, and searching in a web archive (Costa, 2021; Holzmann & Nejdli, 2021; Winters & Prescott, 2019; Jackson et al., 2016; Nielsen, 2016). URL search is offered by most web archives as an entry point to find archived web materials, such as the UK Web Archive, Archive.today, and the Internet Archive’s Wayback Machine, requiring that the user knows the URL in the first instance. Alphabetical browsing is offered by a few web archives such as the UK Government Web Archive and the PRONI Web Archive, while other web archives offer browsing through topical collections, such as the UK Web Archive and the BnF Archives de l’internet (Vlassenroot et al., 2019). Several web archives allow for a full-text search, however, for large web archive collections, this presents challenges due to the huge amount of query returns, which also have a temporal dimension (Winters & Prescott, 2019; Jackson et al., 2016b; Nielsen, 2016; Costa & Silva, 2010). Furthermore, full-text searching within web archive collections “does not provide the same experience of search, or the behaviours of ranking we experience on the live web with search engines such as Google or Bing” (Healy et al., 2022, p. 8; Winters & Prescott, 2019, p. 398). However, as noted by Healy et al. (2022) search capabilities are also a challenge for web archive creators. It is expected in current web design that search boxes or interactive filters are part of navigating a website on the live web. However, these features in the archived version of the website are redundant as the web archive crawler can only follow clickable links and cannot replicate the dynamic interactions that are part of the live website.

Legislation on copyright and legal deposit also presents challenges for researchers to utilise web archives. Using the UK Web Archive legal deposit collections as an example, Winters (2020a) and Milligan (2015) discuss the challenges in using legal deposit collections which are only accessible on a library terminal in a designated reading room. Such challenges include the locked down nature of the library terminal whereby researchers cannot view the source code, and so, it is useless for studies in the evolution of code/CSS design which is important for “web historiography”; nor can a researcher copy the URL from the browser, which causes problems for citation (Winters, 2020a, p. 164; Milligan, 2015). Users are not allowed to copy and paste text which totally disrupts the affordances that are used by researchers worldwide, when they use the live web as a source for research (Milligan, 2015). Also, users can not take photographs or screenshots of the screen, rather they must pay for a printout of an archived web page, which is ironic, as researchers are allowed to use cameras to take photographs of historical documents in most archival environments (Milligan, 2015). Moreover, no two

people can view the same instance of an archived web page simultaneously, even if viewing the same content at different Legal Deposit Libraries, which inhibits collaborative research as well as the use of the resource for teaching in the context of classroom group projects (Winters, 2020a, Talboom, 2022). Such challenges are manifested due to the restrictive nature of the UK legal deposit legislation as laid out in The Legal Deposit Libraries (Non-Print Works) Regulations 2013 (NPLD).

Gooding et al. (2019) offer other examples for the challenges with the NPLD access protocols. For instance, they discuss how the disciplines like digital humanities, data sciences, and quantitative social sciences have evolved to require “libraries to develop new forms of licencing, collection management and support for digital materials in response to user needs” and how the UK government has supported “computational research through a 2014 copyright exception that allows non-commercial text and data mining of copyrighted materials” (p. 7). Nevertheless, they highlight how this sentiment is not extended to NPLD collections, as the NPLD regulations “make no allowance for text and data mining, or to allow materials to be made accessible at the end of their copyright term” (Gooding et al., 2019, p. 7). They suggest that the lack of planning for text and data mining is “now a significant barrier for innovative research” (Gooding et al., 2019, p. 24). Another problem relates to how government reports emphasise “inclusion and access” and how “scholarly publishing is increasingly transitioning towards Open Access” which is also supported by government and research initiatives, and they highlight how “copyright regulations have been enhanced to allow the provision of accessible copies of materials for readers with a recognised disability” (Gooding et al., 2019, p. 7). Yet, aspects such as these

are not formally reflected in the NPLD regulations, which use as a basis the Copyright, Designs and Patents Act 1988 (1988) as amended by the Copyright (Visually Impaired Persons) Act 2002 (2002). This means that the 2013 regulations only allow for accessible copies of NPLD materials to be made available for readers with visual disabilities, rather than all persons with a recognised disability. As such, there is a gap in understanding of the extent to which NPLD supports emerging practices relating to Open Access and accessibility for disabled readers (Gooding et al, 2019, p. 19).

Elsewhere Gooding et al. (2021) suggest that the user was neglected as a stakeholder when it came to drafting the legislation for NPLD access protocols, which is fundamentally print-centric, and because of this, they insist that it fails to consider the user in line with digital user expectations, and current trends in information seeking behaviours (Gooding et al., 2021). Thus, when it comes to evaluating resources like legal deposit collections, in particular the

use of collections with restrictions, it needs to be clearly examined in relation to the rapidity in which technology changes the landscape for end users.

Maurer (2022) also notes how the provision of onsite 'only' access to web archive collections in a designated building makes web archives geographically inaccessible for many researchers (Maurer, 2022). Thus, while some legal deposit schemes might allow for the collection of websites, they may not effectively deal with the provision of access (Healy et al., 2022). Maurer (2022) further suggests that with these types of "closed" archives there is usually very little data

publicly available about their contents, so it is difficult to convince researchers to travel to the reading room when the researcher doesn't know in advance what exactly the archive contains and whether it's pertinent to their research question.

Therefore, one could argue that these types of access conditions present barriers for innovation.

Truter (2021) also highlights the challenges for end user researchers in terms of the access and use of archived web content due to legal restrictions, inclusive of copyright and third-party ownership, privacy policies, and the General Data Protection Regulation (GDPR) in the European Union (EU). This manifests challenges for not only the use of the data, but also affects how and if the data can be made shareable and reusable (Truter, 2021) and runs counter to the requirement of open science which is being stipulated by a growing number of research institutions and funding agencies (Winters, 2020a). However, as pointed out previously, legislation is also a challenge for the creators of web archives, for both its collection, and the conditions for access. Moreover, Healy et al. (2022) highlight how "the circumstances (legal, ethical, curatorial, financial, technical, temporal, social, and political) under which an organisation (or individual) archives web collections" (p. 13), will also affect how end users can access, use, and interpret such collections (Ben-David, 2021; Brügger, 2021c; Ogden, Halford & Carr, 2017; Ogden, 2021; Vlassenroot et al, 2019).

Researchers may also be more interested in using big data methods such as topic modelling or network analysis on a web sphere of websites (WARC files) from a specific web archive collection (e.g., Geocities) or to do a longitudinal study across multiple legal deposit annual web domain collections (see Milligan, 2019; Brügger et al., 2017; Brügger et al., 2019). However, Maurer (2022) points out that "handling the raw WARC files [is] difficult for all parties, so sending extracts of data from institution to research team is often not feasible." Reasons for this are varied and may be "due to a mix of curatorial, technical, legal, economic and organisational constraints" (Brügger, 2021c, p. 217). The Archives Unleashed project

based in Canada, in collaboration with the Internet Archive, sets out to provide some solutions to this issue through the development of tools and solutions for handling and analysing large volumes of WARC files that are hosted on the Archive-It platform. However, this would not apply to heritage institutions who collect web content through in-house crawling for example. This is why Brügger (2021c) stresses the need for solid research infrastructures between the web archives with the data, and the research teams wishing to use the data, and this will help overcome some of the legal, ethical, and technical challenges for both communities. Although, this will require funding, and a cultural shift placing the creator and user as partners in the full web archiving lifecycle.

Challenges for researchers also arise due to ethical issues. Graham (2017) argues that there has been little attention paid to “the ethics of experiencing and accessing the past web” (p. 103). For example, Graham (2019) highlights ethical challenges regarding biases, and reminds us that “on the live web, biases are embedded into both the content and the discovery processes” of what is being collected by web archives. Therefore, Graham (2019) asks how web archivists are “replicating and/or intervening in how biases operate?” once web content is collected and moved “into the more fixed platform of the web archive” (p. 104). Maemura (2018) points to challenges due to “ethical implications of how materials are used”, as well as “questions of consent” and the responsibility of the researcher to the people represented in the data (p. 331). Ogden et al. (2022) suggest that researchers need to be vigilant using web archives when researching socially vulnerable communities and highlight the importance of considering “how particular vulnerabilities can be exacerbated over time when linked to individuals—for example, when researching children, social stigmas (e.g., self-harm communities), or identifying past evidence of illegal activity” (p. 17).

While noting the value of web archives as resources for researching online communities and grass-root histories, Mackinnon (2021) also warns researchers of “significant ethical, methodological and epistemological issues” when it comes to the study of websites of “young people of the past”, and stresses the need for researchers “to consider whose stories are being told, who is equipped to tell them, and what kinds of vulnerability and harm one might encounter and create when doing so” (pp. 442–443).

Other scholars illustrate challenges in using web archives due to political and sociotechnical circumstances. Ben-David (2019) discusses how a ccTLD is delegated to countries which have been recognised by the United Nations (UN), and how “the marked boundaries of these portions of the web comply with the geographical borders that divide nation-states”, and notes how “this assumption is grounded in the practice of web archiving at national libraries” (p. 90). However, for Ben-David (2019) this assumption becomes problematic when it comes

to studying web histories of countries that do not have a ccTLD, such as Kosovo. While Kosovo declared unilateral independence from Serbia in 2008, and was recognised by 133 countries, it is not recognised by the UN due to a veto by Russia, and so, it does not qualify to be allocated a ccTLD (Ben-David, 2019, pp. 91–92). Thus, a Russian veto in the UN for the “recognition of Kosovo has had immense implications on its official presence as a national web”, and consequently presents challenges for tracing Kosovar website histories in a web archive (Ben-David, 2019, p. 95). Ogden and Maemura (2021) examine how the sociotechnical, organisational, and resource constraints “under which most web archiving programmes operate” needs to be understood by researchers, and suggest that researchers need to become familiar with the “specific limits and constraints, legal governance frameworks, collection mandates, as well as configurations (i.e. of sub-collections) and terminology used for specific collections” (Ogden & Maemura, 2021). Schafer et al. (2016) also discuss “the multiple socio-technical mediations, arrangements, and agencies mobilised throughout the archiving process – be they technical or human”, and suggest that an understanding of a web archive “implies opening several black boxes” being that of its collection in order to understand “the human and technological decisions which lead to its constitution, as well as the creation of this source which is never an exact copy of the original” (pp. 2–3).

Elsewhere, Brügger (2016) and Schafer (2019) suggest that web archives present challenges due to the “absence” of a traditional style catalogue or registry as an entry point. Costea (2018) identifies a need for improvements to web archives in the areas of discoverability options, data selection, data management, and access to more comprehensive documentation and metadata. Winters (2017) points out that a major challenge for historians in

working with web archives is, quite simply, that it is difficult; it requires skills that many historians do not have, and in the short term may be unwilling to learn; it involves acknowledging a degree of ignorance with which otherwise seasoned researchers may be uncomfortable (p. 174).

Truman (2016) suggests that challenges arise for researchers due to a lack of technical knowledge in the application of data mining techniques to vast volumes of data, as well as a lack of training and experience in using web archives from discovery processes to integrating the use of archived web content with traditional research approaches. Ruest et al. (2021) also refer to the challenging nature of using big data from web archives due to the “size on the order of petabytes, billions of words, tens of thousands of images, all with murky metadata, provenance, and difficulty to access” (p. 6). Whereas traditional researchers may want to take a more qualitative approach towards using the archived web, they too have challenges due

to a lack of research methods and theoretical paradigms for the use of the archived web (Millward, 2015).

Other challenges relate to the fact that some large-scale web archives, such as the Internet Archive's Wayback Machine, may lack depth and are deemed as too broad to meet the needs of specific research which often requires precise datasets (Schneider et al., 2009, p. 215; Dougherty & van den Heuvel, 2009, pp. 1–5). Therefore, researchers often turn to developing their own web archive collections for their needs (for example, Foot & Schneider, 2006; Engholm, 2000), and so, the research question sets the tone for the identification of what gets captured and why, as well as for how often and for how long. The research question will also drive the approach and methodology, and thus, various methods may be used for the “active process” of archiving web content, such as saving a screenshot or a PDF of a web page, screencasting the functionalities of a web page, or using an archival crawler to collect WARC files (Brügger, 2018; pp. 81–82). For instance, a researcher may collect WARC files to conduct network analysis or topic modelling, or collect images or screenshots for visual discourse analysis, or collect HTML or CSS files to study the evolution of code over time. However, such collections are often narrow in scope and may never be useful for anything other than the study for which they were created (Dougherty & van den Heuvel, 2009, pp. 6–7). Creating a potential for combining researcher-centric web archive collections and derived datasets into collaborative entities may be one answer, but who takes responsibility for the custody, maintenance, and finance of such an entity? Moreover, as different methods are often used for capturing, storage, preservation, and metadata, how would it become interoperable across systems? On the other hand, some researchers do turn to the Wayback Machine for single site histories or to develop their own curated collections and derived datasets for code analysis (e.g., trackers, third-party cookies etc.) or network analysis (e.g., hyperlinks) (Rogers, 2019, pp. 51–53).

Subscription based web archiving services are also another option for researchers to self-archive and curate web archive collections. For example, the Internet Archive's Archive-It service offers a subscription service, based on the amount of data that is archived annually (Archive-It, 2021). Archive-It offers a number of crawling options through this service but for the most part content is archived through Heritrix, or Brozzler which is able to capture more dynamic content. Nevertheless, dynamic or interactive web content, especially social media is still challenging to archive through this service, even with Brozzler. More recent developments in web archiving technologies include the Webrecorder suite of open source tools which enable the user to capture and replay interactive web content as accurately as possible. Webrecorder includes archiveweb.page, a Chrome extension and desktop app



which captures and replays content in a browser and `replayeb.page`, a serverless web and desktop app for viewing web archives directly in the browser. In addition, Conifer is a hosted web archiving service from Rhizome based on Webrecorder software. Webrecorder saves content in either the WARC or WACZ format meaning it can be easily ingested and played back in web archive service infrastructure. Webrecorder software is a largely manual approach that can be accessed on a personal computer making it accessible and easy for any user regardless of technical ability to engage in web archiving. Some institutional web archives use these tools to archive web content deemed important for specific collections. However, due to the time that this manual process to web archiving takes, this cannot be done at scale. The UK Web Archive team at the British Library experimented with using Webrecorder to capture the social media accounts of the main UK political leaders in the 2019 UK General Election. “The major takeaway was that it was a lot more time consuming than we expected. Instead of taking up one working day, it took nearly a whole week to archive our targeted social media accounts with Webrecorder” (Bingham et. al., 2020).

Table 2.1: Useful self-archiving web services for researchers and other users

Provider	Service	URL
Archive.today	Webpage capture	<a href="https://archive.ph/">https://archive.ph/</a>
Arquivo.pt	Save Page Now	<a href="https://arquivo.pt/services/savepagenow">https://arquivo.pt/services/savepagenow</a>
NLI Web Archive	Suggest A Website	<a href="https://www.nli.ie/collections/our-collections/web-archive">https://www.nli.ie/collections/our-collections/web-archive</a>
UK Web Archive	Save a UK website	<a href="https://www.webarchive.org.uk/nominate">https://www.webarchive.org.uk/nominate</a>
Wayback Machine	Save Page Now	<a href="https://archive.org/web/">https://archive.org/web/</a>

For researchers working independently of an institutional web archive there is the additional challenge of storage and preservation of WARC files. Some web archiving initiatives accept nominations from the general public either as part of a specific campaign or as part of their mainstream service. Content (if in scope) can be nominated to the UK Web Archive for preservation, and the NLI Web Archive offers an option to “Suggest a Website” as part of their efforts. Also, sites can be automatically self-archived in the Internet Archive’s Wayback Machine, Archive.today, and in the Portuguese web archive Arquivo.pt (Table 2.1). These are useful services for researchers who want to ensure that the content they are consulting in their research is archived at the same time (or near to that time, in the case of the UK Web Archive) as they are consulting it. In addition, this material can be preserved in more than one

place, making it more accessible to other researchers once the content changes or goes offline (Byrne, 2022).

### 2.3.2 Scholarly use of web archives

While it may seem obvious that historians, media scholars and social scientists will use web archives as resources to document histories of the 1990s and the early millennium, this is not the case (Winters, 2017, p. 174; Ruest et al. 2021, p. 6). Rather, the use of archived web materials for research has only started to see progress in the past decade, or so (Maemura, 2022; Gomes et al., 2021a). This is evident in the volume of edited collections and monographs published on the topic in the last number of years alone (Gomes et al., 2021b; Brügger & Laursen, 2019; Brügger & Milligan, 2018; Brügger & Schroeder, 2017; Brügger, 2017; Milligan, 2019; Brügger, 2018). In addition, there have been a growing number of journal publications, conference papers, and conference presentations which discuss the use of web archives as resources for research, or which offer case studies in the use of web archives and archived content. Such literature is spread across various fields of the academy covering topics such as media and journalism, social sciences and ethnographies, public health and telemedicine, information science and law, internet studies, web histories and more (Schafer et al., 2019; Weber & Napoli, 2018; Bødker & Brügger, 2018; Aust, 2014; Ogden, 2021; Ogden et al., 2022; Gorsky, 2015; Adelman & Franken, 2020; Cocciolo, 2015; Holzmann et al., 2016; Costa & Silva, 2010; Eltgroth, 2009; Taylor, 2017b; Nanni, 2017; Rogers, 2017; Raffal, 2018; Aasman, 2019; Paßmann et al., 2022).

#### 2.3.2.1 Digital and web historiography

Becker (1938) succinctly sums up historiography as “the study of the history of historical study” (p. 20). On one hand, Becker (1938) notes that an objective of “historiography is to assess, in terms of modern standards, the value of historical works” (p. 20). On the other hand, Becker (1938) infers that historiography might be treated

as a phase of intellectual history [...] which records what men have at different times known and believed about the past, the use they have made, in the service of their interests and aspirations, of their knowledge and beliefs, and the underlying presuppositions which have made their knowledge seem to them relevant and their beliefs seem to them true (p. 21).

In explaining web historiography, Brügger (2012) puts forward two distinctions in terms of a web historiography aiming at writing the history of the web, or “a ‘web-minded’ historiography that is a historiography which pays attention to the role of the web in present

day society” (p. 103). In explaining further, Brügger (2012) notes it is worth understanding that since the development of the web, it has become “an integrated part of historiography since more and more documents are made available on the web, and the web is part of the historian’s methodological toolbox” (p. 103). Moreover, the web as a platform also informs “contemporary historiography” as part of a wider movement incorporating humanities and social sciences and the extensions to digital humanities, e-research, and digital research infrastructures (Brügger, 2012, p. 103). Therefore, this allows for a new approach to studying a phenomenon which was not possible prior to the computational turn, the digital turn, or the invention of the web. This would imply that both digital and web historiography needs to be understood in relation to what Brügger (2018) refers to as the “digitality” of the used sources/documents, as well as the methodological processes and tools used to answer the research questions (see [section 2.2](#)).

Winters (2018) further points to the need to understand “the technological contexts” in which digital media has emerged, in the same way that a

palaeographer knows how parchment and ink were produced in the thirteenth century, and so too the digital historian will be required to know how the internet works, and how algorithms are constructed (p. 285).

Moreover, Winters (2018) notes how historians who are accustomed to working with manuscripts will also know how the manuscript was produced and came into being, “and what this means for any analysis of its content” (p. 286). Therefore, for Winters (2018) the “same is true for digital data and the social, cultural, and technical infrastructures which underpin its creation and transmission” (p. 286).

Lay (2017) emphasises how historiography is “one of the essential tools for unlocking the past” and accounts this to an understanding “that history never stands still, that it is argued over and contested.” This will also hold true for digital and web historiography as new methodologies are born out of necessity to deal with the advances of the internet, web and software technologies and the continual evolution of digital media. In the meantime, other methodologies will be doomed due to software incompatibilities or obsolescence and outdated digital media formats. This is not something new, as pointed out earlier, archivists, librarians and information professionals have been discussing this for years. The big question here is how this affects the use of digital materials for academic research whether they are digitised or born digital on the live web, and how can we ensure the use of such digital materials remain accessible, allowing for research reproducibility.

One answer to this is summed up succinctly by Brügger (2018) below:

Studies of the online web have to be documented before, during, or after analysis, to provide a stable object of study and to enable peers to examine the results. Therefore, the question of archiving the web is at the core of an academic study of the web (p. 4).

Thus, we need to consider what it means to study the web relying on both the (born digital) live web and the (reborn digital) archived web as sources for academic study.

### 2.3.2.2 Five strata for studying the archived web

For Brügger (2018), the web can be examined through an analytical grid of five strata: an individual web element, an individual web page, an individual website, a web sphere, the web in its entirety (p. 31). Moreover, studies of the web may have an overlap of strata cases (Brügger, 2018, p. 68). For Brügger (2018) the five strata can be applied equally to both layers of the web being “the visible/audible web in the browser, and hidden text of HTML code and associated files” (p. 31). These five strata also offer an equally applicable model for studying the archived web.

The first stratum is a web element, which Brügger (2018) describes as “any coherent semiotic entity in the form of a written element, a static image element, a moving image element, or a sound element” (p. 33). For example, this may include a banner or an image on a web page, an embedded video, a podcast, footer items, text in the form of a heading, a paragraph, or all text wrapped in a <p> tag in the HTML code. Brügger (2018) offers an example here of Morris’s (2019) study of sound elements from the early web in the Wayback Machine, and Cocciolo’s (2015) study, also using the Wayback Machine, to investigate whether the use of text on web pages was in decline, and if so, by how much.

The next stratum deals with a web page which Brügger (2018) describes as “whatever is presented within a single browser window” (p. 33). So, it is delimited by the “borders” of the window, but Brügger notes that while the word “page” is used it should not be understood in the same way that we think about a page of a book or a print document (Brügger, 2018, p. 34). Rather, one needs to think of it as an interactive document which contains a large element of text, but also contains images, video, audio, maps, databases, and software. Thus, for Brügger (2018) studying web pages may focus on the elements that are presented in the browser window such as text, images, and video, or by analysing “the overall composition of the web page”, for the study of web design evolution for instance (p. 35). Here Brügger offers an example of Richard Rogers’ study of the Google home page from 1998 to 2007 (Rogers, 2013, 2017).

When it comes to describing the website as a stratum, Brügger (2018) proposes it to be

an analytical unit composed of interrelated web pages [which] are connected by semantic, formal, and physical performative means, and the more consistent these three types of interrelations are, the more clearly the website will be delimited (p. 34).

In expanding further, Brügger (2018) suggests that

what delimits the website as such is the extent to which a number of web pages, treat the same subject (semantic cohesion), resemble each other (formal cohesion), and make it possible to go from one page to another (performative cohesion) (p. 34).

Hence, the study of a website might constitute the characteristics of the web pages within the site, or the types of interrelations (Brügger, 2018, p. 34). Research which focuses on individual websites include Hofheinz's (2010) study on the history of *Allah.com* and its creators and influencers, using the Wayback Machine, the live web, and personal archiving. Also, Bødker and Brügger (2018) use a case study of *The Guardian's* website in the Wayback Machine from 1996 to 2015 to examine the shifting temporalities of online news.

The fourth stratum is a web sphere. Foot and Schneider (2006) coined the term web sphere "as a set of dynamically defined, digital resources spanning multiple Web sites deemed relevant or related to a central event, concept, or theme" (p. 20). For example, it might refer to the proliferation of interconnected web content produced because of a natural disaster or be concerned with the web content produced during a sporting event or an election campaign. Thus, it may have a temporal or geographical dimension (Rogers, 2013, p. 74; Webster, 2020, p. 2). Ogden et al. (2022) propose three types of web spheres which are of common interest for study by researchers being "(1) national web domains; (2) platforms, communities, and online ecosystems; and (3) events" (p. 7). Examples here include the work of Brügger et al. (2017) who use the Danish national web archive, Netarkivet, and the Wayback Machine to examine the development of .dk domain names and the .dk domains from 2005 to 2015. Millward (2015) uses the UK Web Archive to "build a corpus of disability websites and pages" through a web sphere of key disability organisations in the UK on the early web (Millward, 2015). Millward then tests a selection of the web material in the corpus with "code validation tools to see whether they conformed to accessibility standards as set out by the World Wide Web Consortium" (Millward, 2015). Beaudouin et al. (2019) offer another example using the French national web archive (BnF Archives de l'internet) to conduct a network analysis "of websites related to the First World War with the aim of understanding how the collective memory of the war is constructed online" (p. 440).

Finally, the fifth stratum is the web in its entirety, although for Webster (2020) “the task of studying the whole Web in terms of its content (rather than its technologies) has so far proved too vast and so has seldom been attempted” (p. 1). However, Brügger suggests that one might consider Weber’s (2019) chapter on ‘Browsers and Browser Wars’, as an example of a study which deals with a part of the history of the whole web. Webster (2020) adds that the fifth strata might also include “studies which illuminate the nature of the whole Web by an analysis of elements that recur across it” (p. 1). Here, Webster (2020) offers an example of Helmond’s study of the “changing purpose, use, and function of the hyperlink over time” (Helmond, 2019, p. 229). Thus, Brügger’s (2018) five strata offer a useful model for combining both the (born digital) live web and the (reborn digital) archived web as sources for academic study and is a useful starting point for the scholarly use of web archives.

## 2.4 Web Archive User Studies

There have been several web archive user studies conducted to date. Some studies focus on engagement with web archives by users in general (Jatowt et al., 2008; Costa and Silva, 2010) or scholarly awareness of web archives, and scholarly research engagement and non-engagement (Hockx-Yu, 2014; Riley & Crookston, 2015; Costea, 2018). Other studies focus on web archiving practices and by correlation how web archive collections are being used (Bailey et al., 2014; Bailey et al. 2016; Farrell et al., 2018), or use a combined focus on web archiving practices and challenges for research engagement (Dougherty et al., 2010; Truman, 2016; Vlassenroot et al., 2019). Healy et al. (2022) examine the challenges for engaging in web archiving, curation, and the use of web archives and archived web content for research or other purposes. Truter (2021) offers one of the few studies which specifically looks at research data management and data sharing practices of researchers in ‘Web Archive Studies’. Here Truter is referring to researchers who use web archives, and archived web data as part of their studies. Outlined next are observations from literature which have some relevance for awareness of web archives, user engagement, and scholarly engagement. Only sources in English are covered as it is the native language of the authors.

In an early study related to access and use of archived web content in Japan, Jatowt et al. (2008) conducted a survey of 1,000 internet users to gain insights on the possible types of interactions participants might have with document histories in a web archive. By document histories, they refer to the different versions of captured web pages. In examining how many participants used web archives in comparison to other web resources, it was revealed that only 1.9% of respondents used a web archive, such as the Internet Archive's Wayback Machine, at least once a month. Jatowt et al. suggest a possible reason for this, may be due

to “the lack of large Web archives open to the public in Japan” and many respondents seemed to be unaware of “the existence of Web archives” (p. 11). In response to the type of information respondents would like to obtain if they could access page histories, 34.2% of participants selected the choice of information about the age of the site, and 21.1% selected the choice for information about the age of the page. In another question, related to access to past content of web pages, participants were asked what they would like to see if they could access the past content of a visited page. Participants were provided with a choice of answers, with two of the top answers being: 49.4% of participants wanted to revisit content that had already disappeared and 29.2% wanted to view content that could not previously be accessed.

Their final question concerns the types of pages for which participants would have liked to view their histories. As a first preference, 42% of participants preferred to view the histories of news sites, and 30.7% preferred to view the histories of pages related to their interests and hobbies. They find that the “ types of pages for which users want to see historical data can vary from person to person” (p. 13), and suggest that “archivists, historians or other professionals may have different requirements and needs regarding the types of documents to be archived” (p. 11). Therefore, they consider that “end users should decide what types of data should be preserved and what types of access should be provided to gain entry to such information in order to make it popular and useful” (p. 13). They also present a case for the possibility of using web archives for comparing historical information with actual information on various real-world objects, such as companies or politicians for example, and suggest that users should “be entrusted with more power to assess the quality and characteristics of real-world objects, be it companies, institutions, or persons” (Jatowt et al., 2008, p. 13).

Examining scholarly engagement, Hockx-Yu (2014) offers a secondary analysis of data that was collected through a user study conducted by the British Library in 2012. The purpose of the user study was to examine the perceptions of scholars for the research value of the Open UK Web Archive, and to gather feedback on access mechanisms; identify gaps in content; and develop a better understanding of the use, or lack of use, of web archives by researchers. The article gives a comprehensive overview of the type of data that the web archive collects and how it was presented to researchers at the time of the British Library study. It further explains the challenges of balancing users’ expectations alongside technical as well as legal limitations. The British Library study found that those who valued the archive the most were scholars interested in web history, statistics, and digital preservation research. From this, Hockx-Yu summarises three approaches to engaging researchers with web archives. The first is in curating thematic collections as a research output, the second is collaborating with

researchers to help them better understand what a web archive is and support them during their research project while the third is the independent use of web archives. Hockx-Yu further discusses the benefits as well as challenges with these three trends (Hockx-Yu, 2014).

On behalf of the National Library of New Zealand (NLNZ), Riley and Crookston (2015) undertook a study via a survey of academics in New Zealand in the disciplines of humanities and social sciences at seven universities and one wānanga tertiary institution (education in a Māori context). The aim of the survey was to gain some insights on the awareness of the existence of web archives by university academics, and to establish more understanding of the use of archived websites by university academics. It further intended to configure what else the NLNZ could do to assist third-level teachers in the provision of access to archived websites for educational benefits. The findings indicate a large lack of awareness by researchers in tertiary institutions in New Zealand. Other findings suggest that respondents preferred a text search option rather than the URL search, which was the only search option available for the NLNZ web archive at the time of the study. Hence, the researchers acknowledge that the access mechanism did not meet the needs of researchers. Respondents also indicated a desire for access to the NLNZ domain harvest via full text search and demonstrated a requirement for the NLNZ to develop options for making this available. Finally, 51% of researchers in the study indicated that the New Zealand Web Archive will become important for their research within the next five years (Riley and Crookston, 2015).

In Denmark, Costea (2018) conducted a study with the aims of providing some perspectives on researcher engagement with web archives, researcher needs in the use of web archives, and to identify reasons for the non-use of web archives by researchers. The study targeted professors, researchers, and PhD students from the Arts, Humanities, and Social Sciences in two Danish universities. It used a mixed method approach of an online survey (n=88), semi structured interviews (n=3) and testing with first-time users (n=2). After analysis, Costea found a noteworthy lack of awareness of the existence of web archives as resources for research. Many researchers were unaware of the content of a web archive, and how a web archive can be used as a resource for research. Users and non-users alike appreciated the value of archived web content, but also identified a need for improvements to web archives to satisfy researchers' needs in the areas of discoverability options, data selection, data management, and more access to methods for data analysis. Issues of incompleteness of data in web archives were also mentioned. Access to more comprehensive documentation and metadata was thus seen as a requirement for researchers. Findings from both the survey and interview also highlight the need for researchers to be able to extract data from a web archive to create a dataset for their own research needs (Costea, 2018).



Truter's (2021) study is one of the few studies which examine research data management and data sharing practices of researchers who use web archives as part of their studies. Truter's study combined a survey targeted at international 'Web Archive Studies' researchers (n=31), and one semi-structured interview with an individual with experience of working with research data from web archives. Truter found that the main challenge for sharing archived web data/materials is legal restrictions, inclusive of copyright and third-party ownership, privacy policies, and GDPR, which creates challenges not only for the use of data from web archives but may also affect the ability to share the data or make it reusable. This study further highlights challenges with the volume of data as well as the complexities of different media types and formats. The lack of a dedicated repository for the long-term preservation of archived web data; difficulty with Data Management Plans (DMPs); and a lack of storage space were also highlighted as significant challenges by participants. Other challenges included a lack of funding for research data management, and a lack of guidance/training provided by publishers for those undertaking research in web archive studies (Truter, 2021).

The National Digital Stewardship Alliance (NDSA) undertook surveys of web archiving initiatives in the United States in 2011, 2013, 2016, and 2017. For the most part the aim of each study was to explore the types of web archiving programmes being conducted, the background and extent of their web archiving activities, the types of content they select for preservation, the types of tools and services they use, the types of access and discovery options they provide, the types of permissions they seek for collection and access, and the types of policies they have in place. The NDSA was founded in 2010 and is made up of educational, governmental, non-profit, and commercial organisations, who have a vested interest in the long-term preservation of digital information (Farrell et al., 2018, p. 4). While its original membership was confined to organisations in the US, it was extended to international organisations in recent years, and at the time of writing it comprises more than 260 members (National Digital Stewardship Alliance, About Us, n.d.). Of interest for the current research are their observations on the use of web archive collections by researchers.

The first study conducted in 2011 study finds "an area of uncertainty" by web archiving institutions vis-à-vis how collections were being used (NDSA Content Working Group, 2012, p. 10). For example, in response to a question on how researchers are using their archive, a large majority responded with variants of "unknown"; "too soon to tell"; or "good question" (NDSA Content Working Group, 2012, p. 11). The report observes that "the lack of knowledge about web archive usage and users" is clearly a topic that merits further investigation" (NDSA Content Working Group, 2012, p. 11). The survey conducted in 2013 does not appear to have asked a question on user engagement, and the question on how researchers are using web

archives does not appear to have been asked in the questionnaire. Regarding whether they had active researchers utilising their web archive, in the survey conducted in 2016, there were 80 responses of which 19% (=15) answered 'Yes', 30% (=24) answered 'No', and 51% (=41) answered 'Don't know' (Bailey, et al., 2017, p. 27). Again, the study highlights the lack of understanding in how collections are being used as "an area of activity that merits community attention" (Bailey, et al., 2017, p. 27). In the 2017 study, the same question regarding the use of their web archives by researchers had 117 respondents, of which 18% answered 'Yes', 33% answered 'No', and 49% answered 'Don't Know' (Farrell et al., 2018, pp. 23–24). While this question specifically targets use by researchers, it provides some indications of the extent to which there is a lack of awareness by web archiving institutions apropos how their web archives are being used.

Finally, the WARST project conducted a study which examines the skills, tools, and knowledge ecologies in web archive research (Healy et al., 2022). WARST is a collaborative project by researchers from Maynooth University, the British Library, the International Internet Preservation Consortium, the Bavarian State Library, and the University of Siegen. The research team are all members of Web ARChive studies network researching web domains and events (WARCnet, warcnet.eu). The study focuses on individuals (N=44) around the globe who participate in web archive research, in the context of web archiving, curation, and the use of web archives and archived web content for research or other purposes. The methodology for the study entailed desk research, participation in WARCnet meeting discussions, and an online questionnaire. The study sought to identify and document the skills, tools, and knowledge required to achieve a broad range of goals within the web archiving lifecycle. It further sought to explore the challenges for participation in web archive research, and the overlaps of such challenges across communities of practice.

The authors organise the findings around two communities of practice being individuals from a 'Library, Archive, or Web Archive environment' and individuals who identified with being a 'Scholar, Academic, Lecturer, Student, or working in an IT/Web Design environment', however for the purpose of this report we will solely focus on the second community. Participants who identified with 'Scholar, Academic, Lecturer, Student, or IT/Web Design environment' (n=9) mentioned challenges related to the inconsistency and incompleteness of the data, a lack of documentation and metadata, challenges which arise due to legalities on access, use, and storage. Participants also found difficulties in the use of web archives due to a lack of research methods and approaches, working with large volumes of data, as well as challenges with learning new skills (Healy et al., 2022). Healy et al. (2022) further discuss how some of the challenges mentioned above overlap between creators and users. For example,

both creators and users have challenges in the areas of search and retrievability, users find it difficult to search large-scale web archives, while creators find it difficult to provide search mechanisms and algorithms for large-scale collections that will satisfy a diversity of users. In addition, both creators and users have challenges with legal issues such as copyright and legal deposit, users have challenges accessing content, while creators have challenges for the collection of web content, as well as the provision of access, and how restrictive this access might be (Healy et al., 2022).

### 3. AVAILABILITY AND ACCESSIBILITY OF IRISH BASED WEB ARCHIVES

*Finding a balance between preservation and access is the most urgent problem to be solved, because if today's Web is not saved it will not exist in the future. Access is a political as well as a legal problem. The answer to the access problem, like the answers to all political problems, lies in establishing a process of negotiation among interested parties. Who are the stakeholders, and what are the stakes, in building a Web archive? (Lyman, 2002, p. 40).*

#### 3.1 Introduction

In the previous section the literature identified some of the main causes for the loss of digital heritage and explored the relationship between legal deposit legislation and the preservation of national heritage. As part of this, it observed how web archiving is an important activity for the preservation of national digital heritage for future generations and explored some of the challenges for the web archiving community. The section also discussed the challenges experienced by the web archive user community and examined the literature on scholarly engagement with web archives, through a review of user case studies and web archive user studies. This is a useful starting point when it comes to examining the current state of web archive research in Ireland, and how useful web archives are as resources for conducting Irish based research, which has been relatively understudied.

As mentioned before, to the best of our knowledge there have been no studies conducted which offer an overview of the availability and accessibility of web archives based across the island of Ireland as resources for research, which would be useful for mapping the opportunities for the use of these resources for research on Irish based topics. It would also be useful for assessing support and incentive mechanisms to foster developments in the use of web archives for Irish based research and encourage engagement by other types of end users. Therefore, this section seeks to explore the availability and accessibility of web archives which would prove useful for conducting Irish based research.

At present, there are three main web archiving initiatives which capture websites as part of their efforts for the preservation of digital heritage for the island of Ireland. The Public Record Office of Northern Ireland (PRONI) takes responsibility for capturing websites related to the six counties of Northern Ireland through a selective collection approach and provides online

access to their collections via the PRONI Web Archive.<sup>6</sup> The UK Web Archive also has responsibility for capturing websites in Northern Ireland through their annual national web domain crawl, under The Legal Deposit Libraries (Non-Print Works) Regulations 2013 (NPLD).<sup>7</sup> To note here, as a UK Web Archive partner, the Library of Trinity College Dublin (TCD) provides onsite access to the UK Web Archive's NPLD national web domain collections through their onsite library pc system (Library of TCD, n.d., Electronic Legal Deposit). The National Library of Ireland (NLI) takes responsibility for capturing websites as part of their strategy for preserving the digital heritage of the twenty-six counties in the Republic of Ireland. First, through a selective collection approach which is accessible online through the NLI Web Archive.<sup>8</sup> The NLI Web Archive have also conducted two national web domain crawls in 2007 and 2017, however, these collections are currently inaccessible to researchers or the public due to legislative matters and will be discussed in more detail further on.

There is no doubt that Irish digital heritage can be found in other web archives such as the Internet Archive (Wayback Machine), Common Crawl, and Archive.today.<sup>9</sup> Other web archives which have relevance to Northern Ireland are the UK Government Web Archive, and the UK Parliament Web Archive. The UK Parliament Web Archive captures, preserves and makes accessible "UK Parliament information" that is published on the web, and "includes UK Parliament websites and social media dating from 2009 to the present" (UK Parliament Web Archive, n.d.).<sup>10</sup> Also, the UK Government Web Archive captures, preserves, and makes accessible "UK central government information" that is published on the web, and includes "videos, tweets, images and websites dating from 1996 to the present day" (The National Archive, n.d., UK Government Web Archive).<sup>11</sup>

There may also be other minor web archiving initiatives being conducted which have relevance for Irish heritage, by researchers in an academic setting, or for business purposes. For example, the Library of University College Dublin (UCD) collected circa 150 websites "relevant to Irish poetry in the 21st century" which are hosted, and openly accessible on the

---

<sup>6</sup> PRONI Web Archive, <https://webarchive.proni.gov.uk/#!/>

<sup>7</sup> UK Web Archive, <https://www.webarchive.org.uk/ukwa>

<sup>8</sup> NLI Web Archive, <https://archive-it.org/home/nli>

<sup>9</sup> Wayback Machine (Internet Archive), <https://archive.org/web>; Common Crawl, <https://commoncrawl.org>; Archive.today, <https://archive.ph>

<sup>10</sup> UK Parliament Web Archive, <http://webarchive.parliament.uk/atoz>

<sup>11</sup> UK Government Web Archive, <https://www.nationalarchives.gov.uk/webarchive/>

Archive-It service platform.<sup>12</sup> As the collection strategy for the library's Irish Poetry Reading Archive has become increasingly digital, with "a huge amount of material of importance" only being available on the web, they embarked on a project to collect and preserve websites relevant to Irish poetry for future generations (UCD Collections, Archive-IT). Other examples include social media archive collections.

Such archives include a collection by Darcy et al. (2021) in the Digital Repository of Ireland, titled 'In Her Shoes: Stories of the Eighth Amendment', which was collected as one part of a wider collection programme, by the award winning Archiving Reproductive Health project (Digital Repository of Ireland, 2022).<sup>13</sup> The collection comprises of administrative posts and 'stories' which were published on the Facebook page, 'In Her Shoes - Women Of The Eighth', during the run up to the Referendum to the Repeal the Eighth Amendment, colloquially known as the 'abortion' referendum which took place in the Republic of Ireland on 25 May 2018.<sup>14</sup> The referendum was successful, resulting in a constitutional change through the Thirty-sixth Amendment of the Constitution Act 2018.<sup>15</sup> There are also other archived social media datasets which relate to the Repeal the Eighth referendum. These include Littman's (2018) 'Ireland 8th Tweet Ids', which was collected from the Twitter filter stream API using Social Feed Manager and is available in the Harvard Dataverse repository. Also, Ó Briain and Foster's (2020) '#retweetthe8th: twitter dataset' which was collected from the Twitter filter stream API using Twarc, between 09 March and 30 May 2018, and is available in the Zenodo repository.<sup>16</sup>

For this report, however, we are interested in web archive initiatives that are based on the island of Ireland which have a specific mandate to capture a wide range of Irish digital heritage as part of their collection development strategies. Therefore, we focus on the PRONI Web Archive, the NLI Web Archive, and the UK Web Archive, which is accessible onsite in the Library of TCD; and we examine the availability and accessibility of these resources for

---

<sup>12</sup> UCD Special Collections, Archive-It, <https://archive-it.org/organizations/1846>

<sup>13</sup> DRI, In Her Shoes: Stories of the Eighth Amendment, <https://repository.dri.ie/catalog/wm11nd02p>

<sup>14</sup> Eighth Amendment of the Constitution Act, 1983, <https://www.irishstatutebook.ie/eli/1983/ca/8/enacted/en/html?q=Eighth+Amendment+of+the>

<sup>15</sup> Thirty-sixth Amendment of the Constitution Act 2018, <https://www.irishstatutebook.ie/eli/2018/ca/36/enacted/en/html>

<sup>16</sup> Ireland 8th Tweet Ids, <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/PYCLPE;> #retweetthe8th: twitter dataset from the 2018 Referendum to repeal the 8th Amendment of the Constitution of Ireland, [https://zenodo.org/record/3842013#.YZOx\\_S2l1bV](https://zenodo.org/record/3842013#.YZOx_S2l1bV)

conducting research on Irish based topics. In doing so, we offer an overview of these web archiving initiatives and their historical backgrounds, inclusive of how copyright and legal deposit has influenced their collecting activities. We further observe their efforts for the collection and preservation of digital heritage from the web spaces of Northern Ireland (NI) and the Republic of Ireland (ROI), and we assess their availability and accessibility as resources for conducting Irish based research. As mentioned previously, this chapter refers to Irish digital heritage in the context of the island of Ireland, but it is also worthwhile examining some of the historical context in which the heritage of Ireland was preserved preceding the digital turn.

## 3.2 Preservation of Irish Records and Publications, pre-Digital

Prior to the Irish War of Independence (1919-21), the Anglo-Irish Treaty (1921) and the 1920-21 partition of the island of Ireland, under UK legislation the NLI was a nominated institution for the collection of books and publications, while the State Paper Office and the Public Record Office of Ireland were the nominated institutions for the preservation of the state records of Ireland. Following partition, PRONI was established “for the reception and preservation of public records appertaining to Northern Ireland” (Section 1: Public Records Act (Northern Ireland), 1923). In the next section, we briefly look at some of the backgrounds of these institutions.

The NLI was established in 1877, following negotiations between the Royal Dublin Society, the Department of Science and Art (London) and the Commissioners of Public Works (Ireland), which led to the Dublin Science and Art Museum Act of 1877 and the establishment of a national library and national museum. As Ireland was part of the UK at the time, the newly established NLI was governed from London up until the establishment of the Irish independent state in 1922, when it was handed over to the Irish Government under the remit of the Department of Education. From July 1986, the NLI was transferred to the Department of An Taoiseach, and transferred again in 1992 to the newly established Department of Arts, Culture and the Gaeltacht (NLI, n.d., History of the Library).

The National Cultural Institutions Act, 1997 confirmed the NLI as the official library of record in the ROI in its responsibility for collecting for, and on behalf of the Irish state (NLI, Collection Development Policy 2022-2026, p. 2).<sup>17</sup> In 2005, the NLI became an autonomous agency,

---

<sup>17</sup> National Cultural Institutions Act, 1997,  
<https://www.irishstatutebook.ie/eli/1997/act/11/enacted/en/html>

governed by a Board, and is currently under the aegis of the Department of Tourism, Culture, Arts, Gaeltacht, Sport and Media (previously called the Department of Culture, Heritage and the Gaeltacht) (NLI, n.d., History of the Library; Collins, 2018).

Today, the mission of the NLI is to “collect, protect and make accessible the recorded memory of Ireland” inclusive of “books, serials, newspapers, manuscripts, maps, photographs, official publications, prints, drawings, ephemera, digitised and born-digital collections” (NLI, Collection Development Policy 2022-2026, p. 1). In doing so, they are also committed to collecting “representatively and inclusively” in order “to capture the diversity of Irish experience” (NLI, Collection Development Policy 2022-2026, p. 2) and “create a more diverse and inclusive story of Ireland, so that new voices are collected and shared with the world” and ensure “that Ireland is represented in all its diversity, in all of our activities and that equal access to these is provided for everybody” (NLI, Diversity and Inclusion Policy 2018-2021, p. 1). Thus, from 1887 to the current day, the NLI has been a leading force in the preservation of Irish heritage.

In terms of state records, a State Paper Office (SPO) for Ireland was established in 1702, to preserve the records of departing chief governors, and the keeping of “any Kings' or Queens' Letters, warrants, orders, petitions and other letters belonging to the Secretaries' Offices” (Wood, 1930, p. 23). Administered from Dublin Castle, the SPO was handed over to the new independent Irish State in January 1922 (Maguire, 2022) and would continue to operate until the 1980s. The Public Records (Ireland) Act, 1867 provided for the establishment of a Public Record Office of Ireland (PROI) in a purpose-built facility in the Four Courts complex in Dublin, consisting of a “Record House” and a “Record Treasury” (Wood, 1930, pp. 30–33).<sup>18</sup>

According to the National Archives of Ireland website, the buildings

consisted of a three-storey over-basement Record House with staff offices, a caretaker's apartment, a library, a binding room and a public reading room. Behind the Record House was the Record Treasury, an enormous six-storey building containing 100,000 square feet of shelving with records accumulated over seven centuries (National Archives of Ireland, n.d., Public Record Office of Ireland).

The transfer of records to the new PROI facility included records from Dublin Castle, the Landed Estates Record Office at the Custom House, and legal records held by the courts in the Four Courts complex (Wood, 1930, pp. 29–31). Moreover, “Every year the documents in

---

<sup>18</sup> Public Records (Ireland) Act, 1867, <https://www.nationalarchives.ie/PDF/PROI1867.pdf>



the various public offices throughout Ireland which had arrived at twenty years of age were automatically transferred to the Record Office” (Wood, 1930, p. 33). In June 1922, the Record Treasury of the PROI was destroyed by fire and explosion in the opening days of the Irish Civil War (Wood, 1930, p. 35). Regan (2016) posits that the destruction of the records of the PROI marked “a cultural atrocity unique in modern Irish history” (p. 11). By 1928, the PROI was again open to the public, and “herculean efforts were made by the staff to find replacements for records that had been destroyed” (Crowe, 2012). These efforts were further assisted by the establishment of the Irish Manuscript Commission in 1928, who were given the remit of reporting on

the nature, extent, and importance of existing collections of manuscripts and papers of literary, historical and general interest relating to Ireland, and on the places in which such manuscripts and papers are deposited, and to advise as to the steps which should be taken for the preservation and publication of such manuscripts and papers, whether in public collections or in private ownership (Dáil Éireann, Ceisteanna—Questions. Oral Answers - Manuscripts Commission, 17 October 1928; Irish Manuscripts Commission, n.d., Mission).<sup>19</sup>

The PROI and SPO continued to function until the enactment of the National Archives Act, 1986, which established the National Archives of Ireland on 1 June 1988. The Act transferred both the functions and holdings of the PROI and the SPO to the newly established National Archives.<sup>20</sup> Furthermore, “Under this legislation, records of Government Departments and their agencies are transferred to the National Archives when they are 30 years old” (NAI, n.d., About the National Archives). The thirty-year rule will be gradually decreased to a twenty-year rule in the foreseeable future (McGee, 2018).

In 1989, the NAI were assigned new premises in Bishop Street, Dublin, and the SPO in Dublin Castle was vacated in August 1991, and the PROI facility in the Four Courts was vacated in September 1992, as the NAI began operations from their new headquarters in Bishop Street (NAI, n.d., About the National Archives). The “salved records” which had been saved from the rubble of the PROI Record Treasury were finally dealt with as part of an innovative project titled ‘Beyond 2022: Virtual Record Treasury of Ireland’. According to the project’s website,

Beyond 2022 is an all-island and international collaborative research project working to create a virtual reconstruction of the Public Record Office of Ireland,

---

<sup>19</sup> Dáil Éireann. (1928). Ceisteanna—Questions. Oral Answers. - Manuscripts Commission, <https://www.oireachtas.ie/en/debates/debate/dail/1928-10-17/3>

<sup>20</sup> National Archives Act, 1986, <https://www.irishstatutebook.ie/eli/1986/act/11/enacted/en/html?q=National+Archives+Act>

which was destroyed in the opening engagement of the Civil War [...] Together with our 5 Core Archival Partners and over 40 other Participating Institutions in Ireland, Britain and the USA, we are working to recover what was lost in that terrible fire one hundred years ago (Beyond 2022, n.d., Home)

Through the identification of duplicate documents in archives elsewhere, or documents which may reference such documents, and the conservation of the “salved records” from the debris, the project was able to digitise, transcribe, and assign metadata to a rich assortment of replacement materials. Coinciding with the centenary of the Four Courts catastrophe at the end of June 2022, the Beyond 2022: Virtual Record Treasury of Ireland was launched online, bringing together the replacement materials within an immersive 3-D (re)construction of the destroyed Record Treasury building (Figure 3.1). The project was funded by the Irish Government under Project Ireland 2040, through the Department of Tourism, Culture, Arts, Gaeltacht, Sport and Media (Beyond 2022, n.d., Support).



Figure 3.1: Screenshot of the interface of the 3D Virtual Record Treasury of Ireland (<https://vrtour.virtualtreasury.ie>), taken on 2022-10-18<sup>21</sup>

The Beyond 2022 project is a great example of how to create academic and popular awareness of a heritage problem while using technology to provide solutions. The project’s efforts to identify paper documents from multiple archives and libraries, digitise such

---

<sup>21</sup> Beyond 2022, Virtual Tour: Record Treasury of Ireland, <https://vrtour.virtualtreasury.ie/>

documents and assemble such documents as replacement documents for those that were destroyed in the PROI, offers a model which could be used in other countries where archived collections have been destroyed due to fire, flooding or natural disasters for example. However, it should be stressed that it would not be much use as a model for born digital heritage such as websites, as web content often “replaces its antecedent, usually leaving no trace of the previous document/edition”, and once a web document is “permanently removed from the WWW it ceases to exist” (Koehler, 1999). Thus, unless it is archived, it is lost to perpetuity.

While the NAI was essentially established in the digital age, it has had significant challenges when it comes to digital, as outlined in [section 2.1](#). From at least 1997, the NAI have continually warned the Irish government regarding the loss of Irish digital heritage, through the loss of electronic records due to obsolete technology, or the fact that there is a lack of formal record keeping guidelines for electronic information by government departments and agencies. Section 2.1 also demonstrated how content on the Irish government website(s) has changed and disappeared over the past decades. However, the NAI is not currently involved in a public web archiving initiative for the websites of the government, its departments, or associated agencies. While the NAI produced an ambitious strategy in 2021 to deal with the information age including “a digital transformation programme [and] a new framework for records management across government” (NAI, n.d., News; NAI, 2021b), section 2 highlighted that the strategy will depend on “improved funding, an enhanced infrastructure and [...] improved staffing resources (NAI, 2021b, p. 6). As the NAI gradually moves from a thirty-year rule to a twenty-year rule for the transfer of government records to the NAI (McGee, 2018), it is already facing the problem of curating the first wave of digital records from government departments, which is bound to increase exponentially.

Regarding NI, following partition, the Public Record Office of Northern Ireland (PRONI) was established through the Public Records Act (Northern Ireland), 1923. As its purpose it would serve as an archive “for the reception and preservation of public records appertaining to Northern Ireland which otherwise would be deposited in the Public Record Office of Ireland” (Section 1: Public Records Act (Northern Ireland), 1923). The Act also permitted PRONI to collect documents which did not expressly relate to Government creation or use (PRONI, 2008). PRONI combined the functions of a public and a state records office.

PRONI began operations in March 1924 in a disused linen warehouse in Murray Street, Belfast. Under the guidance of the first Deputy Keeper, Dr D.A. Chart, their first mission was to find and recover surrogates of records which had been destroyed by fire and explosion at the Public Record Office of Ireland in Dublin during the Irish Civil War (PRONI, 2007; PRONI,

2008). Having previously worked at the Public Record Office of Ireland in Dublin, Dr Chart was familiar with the records which had been destroyed, and immediately set out to solicit duplicate records held by churches, solicitors, politicians, businesses, and the landed aristocracy – which proved to be very fruitful (PRONI, 2007; PRONI 2008). In April 1933, PRONI moved to the newly built Courts of Justice building, in Chichester Street, Belfast, and moved again in 1968 to a purpose-built location in Balmoral Avenue, Belfast (PRONI, 2007). PRONI moved once more in 2011 to its current location, at a purpose-built facility in the Titanic Quarter of Belfast (NIdirect, n.d., Getting to PRONI).

From 1924, PRONI came under the jurisdiction of the Ministry (later Department) of Finance and moved to the jurisdiction of the Department of the Environment in 1982 where it became an executive agency within that department. From 1999, PRONI became an agency under the Department of Culture, Arts and Leisure with the restoration of a devolved government. From 2006, PRONI ceased to be an agency, and became a division within the main Department of Culture, Arts and Leisure (PRONI, 2008).

Mostly dating from the seventeenth century to the present day, PRONI contains millions of records “that relate chiefly, but by no means exclusively, to Northern Ireland” (CAIN, n.d.). Their holdings fall under the following categories:

- privately deposited archives: e.g., landed estate archives, business records, church registers, emigrant letters, etc.
- public records: e.g., records from official sources such as local authorities, courts of law, quangos, public bodies, etc.
- departmental records of the various departments and ministries involved in the governance of Northern Ireland since 1921 to present (CAIN, n.d.).

## 3.3 NI Web Space

### 3.3.1 PRONI web archive

In NI, PRONI began a selective web archiving programme circa 2010, to capture and preserve websites of government departments, local councils, public sector organisations and websites “of social, political, cultural, religious or economic significance and relevance to Northern Ireland” (NIdirect, 2015; Murchan, 2020a). The resource is publicly available online as the PRONI Web Archive. On the technical side, PRONI partners with Archive-It, a subscription-based web archiving service provided by the Internet Archive in the United States (US). However, they originally partnered with the Internet Memory Foundation (IMF), a

subscription-based web archiving service in Europe, which ceased operations circa August 2018 (Nidirect, 2015; Wikipedia, 2012+, Internet Memory Foundation; Aturban, 2019b).

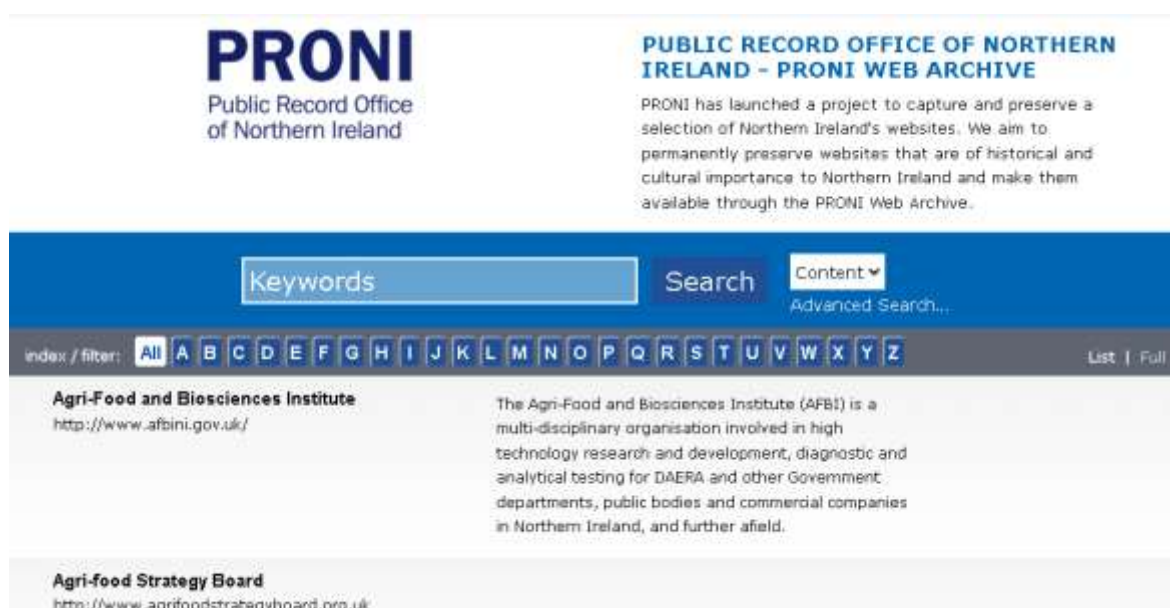


Figure 3.2: Screenshot of the interface of the PRONI Web Archive, taken on 2022-08-24

PRONI operates a two-fold approach for their selective based collection efforts. First, regarding websites under the jurisdiction of NI government departments, arm's-length bodies, or other publicly funded bodies, PRONI “does not seek permission to crawl and publish” such websites, rather it notifies the website owners/government departments of their intentions to include the website in their collection and provides a takedown policy (Murchan, 2020b; PRONI, 2018, p. 5). The takedown policy also accounts for cases whereby a website, or parts of a website, may be in dispute due to third party copyright (e.g., copyrighted photographs), or in breach of data protection due to the availability of an individual’s personal information, for example (PRONI, 2018, p. 5; PRONI, 2016). With the second approach, PRONI does, however, seek permissions “to crawl and publish privately funded” websites which are deemed historically or culturally important for inclusion in the PRONI Web Archive (PRONI, 2018, p. 5). In August 2022 there were 320 websites listed as available, of which many have accompanying descriptive metadata on the interface (see [Figure 3.2](#), [Figure 3.3](#)). The resource can be searched by inputting a known URL, and through a search of the content with keywords. There is also an advanced search option which includes full-text or phrase search and refining the search through date parameters.

<b>Behind the Hill Exhibition</b> <a href="https://www.behindthehill.org/">https://www.behindthehill.org/</a>	After exploring the collections at the Public Record Office of Northern Ireland (PRONI), these artists present new artworks made in response to their research. The exhibition includes a curated collection of items from their research including photographs, diary entries, a letter and a map.
<b>Belfast City Council</b> <a href="http://www.belfastcity.gov.uk/">http://www.belfastcity.gov.uk/</a>	Official website for Belfast City Council
<b>Belfast Education and Library Board</b> <a href="http://www.belb.org.uk/">http://www.belb.org.uk/</a>	
<b>Belfast Festival at Queen's</b> <a href="http://www.belfastfestival.com/">http://www.belfastfestival.com/</a>	
<b>Belfast Harbour Commissioners</b> <a href="https://www.belfast-harbour.co.uk/">https://www.belfast-harbour.co.uk/</a>	Website dedicated to Belfast Harbour noting its history, corporate information, port information, real estate, harbour estate access, news and key documents.

Figure 3.3: Screenshot of the interface of the PRONI Web Archive, showing descriptive metadata entries, taken on 2022-09-26

### 3.3.2 UK web archive

In addition to this, UK NPLD allows for the archiving of the NI web space under the jurisdiction of the UK legal deposit libraries. Access to the NPLD web archive domain collection is only available onsite in a legal deposit library. These include the British Library, the National Library of Scotland, the National Library of Wales, the Bodleian Library at Oxford University, Cambridge University Library, and the Library of Trinity College Dublin (The Bodleian Libraries, n.d., Legal Deposit).

Legal deposit for print publications first became a part of English law through the Licensing of the Press Act (1662) and became a more formalised British law through the Copyright Act, 1710, also known as the Statute of Anne (Partridge, 1938, p. 33; Koehler, 2015, pp. 149–150). The Copyright Act, 1710 formalised a relationship between copyright and legal deposit as part of the same legislation. However, legal deposit was only officially extended to Ireland with the enactment of the Act of Union of Great Britain and Ireland in January 1801.<sup>22</sup> This also formalised the relationship of the Library of TCD as a UK legal deposit library up to the current day (Library of TCD, n.d., Electronic Legal Deposit; Partridge, 1938, p. 45). Partridge (1938) suggests that it was a great failure on the part of the Westminster Parliament not to include

---

<sup>22</sup> Act of Union (Ireland) 1800, <https://www.legislation.gov.uk/aip/Geo3/40/38/contents>

Ireland in the Copyright Act 1710, as Ireland became a haven for piracy with the reprinting of English and Scottish books throughout the 1800s (p. 134).

During the 1800s there were several more changes to legal deposit with the Copyright Acts of 1814 and 1836 which first increased and then decreased the number of nominated legal deposit libraries (Muir, 2005, p. 14). Thereafter the 'Imperial' Copyright Act of 1842 was introduced as an attempt to regulate copyright legislation throughout the British empire (Partridge, 1938, p. 80). Partridge (1938) notes that the Act of 1842 was designed to ensure that the British Museum (later renamed British Library) secured a copy of "every book anywhere under British rule" (p. 80), and copyright law then remained practically unchanged for much of the century (Feather, 1994, p. 6). The Copyright Acts of 1814, 1836 and 1842 also sought to recognise authors, who had up to this point been somewhat neglected by the Copyright Act, 1710 (Feather, 1994, p. 6).

In the 1900s, the Copyright Act 1911 extended the copyright term of an author and incorporated the National Library of Wales as a legal deposit library (Partridge, 1938, p. 108), although it could only claim "material in Welsh or of Welsh or Celtic interest" which would not change until the 1970s (Feather, 1994, pp. 120–121). The 1911 Act stipulated that a copy of print books be deposited, free of charge, in the British Museum within a month of publication, and the other five legal deposit libraries could claim a copy within 12 months of publication (Muir, 2005, p. 19). The five legal deposit libraries included the National Libraries of Scotland and Wales, and the University libraries of Oxford, Cambridge, and Trinity College Dublin (Working Party on Legal Deposit, 1998). Thereafter, the Copyright (British Museum) Act 1915, reduced the types of publications to be deposited in the British Museum, exempting such items as rail timetables, advertisements, voters rolls, design specifications, and calendars (Muir, 2005, p. 18).

Hence, from 1801 up until the establishment of the independent Irish State in 1922, Ireland came under the jurisdiction of UK copyright and legal deposit laws. Thereafter, they only applied to the six counties of NI. It would then take until 1927 for the Irish State to introduce its own legislation on copyright and legal deposit through the Industrial and Commercial Property (Protection) Act, 1927.<sup>23</sup>

There were further changes to UK copyright and legal deposit laws such as The Theatres Act 1968 which allowed for the incorporation of published theatre production scripts as part of

---

<sup>23</sup> Industrial and Commercial Property (Protection) Act, 1927,  
[https://www.bailii.org/ie/legis/num\\_act/1927/0016.html#zza16y1927](https://www.bailii.org/ie/legis/num_act/1927/0016.html#zza16y1927)

legal deposit, although there was a failed attempt to extend legal deposit to films through the Film (Statutory Deposit) Bill in 1969 (Muir, 2005, p. 18).<sup>24</sup> Also, copyright was amended in the Copyright Act, 1956, but it did not affect legal deposit (Muir, 2005, p. 20). Muir (2005) suggests that for legal deposit the Copyright Act, 1911 remained relatively unchanged until the end of the century (p. 23).<sup>25</sup> Nonetheless the inclusion of non-print material cropped up on the agenda from time to time. Muir (2005) notes how several proposals were put forward to extend legal deposit to non-print material such as “microfilm, sound and audiovisual material, such as films” (p. 20). In addition, as the UK was a member of the European Economic Community (EEC) since January 1973, which was later renamed the European Community (EC), and thereafter the European Union (EU), the UK was often subject to directives which attempted to harmonise copyright across member states (The National Archives, n.d., The EEC). For example, EEC member states had to comply with a 1993 Council Directive, which was an attempt to harmonise copyright across member states on account of the

differences between the national laws governing the terms of protection of copyright and related rights, which are liable to impede the free movement of goods and freedom to provide services, and to distort competition in the common market (Council Directive No. 93/98/EEC (2)).

In terms of legal deposit however, it was not until the mid-1990s that legal deposit legislation would really come under scrutiny, due to the “development of new media and the growth of publication in non-print forms” (Working Party on Legal Deposit, 1998). The growth of born digital media could not be denied, due to the rapid development of the web, and the increasing availability of the internet (Masanès, 2006, p. 3).

A proposal to the UK Government to extend legal deposit to non-print materials including digital material was submitted by the British Library in 1996, on behalf of the legal deposit libraries and the British Film Institute. This kick-started a commitment by the UK Government to develop a legal deposit scheme for non-print, with the setup of an interim voluntary scheme for microfilm in 2000, pending forthcoming legislation (Muir, 2005, p. 4; p. 29). In taking the UK into the digital information age, the Legal Deposit Libraries Act 2003 extended the UK legal deposit scheme to “non-print (electronic) publications, including websites, subject to further enabling Regulations” (UK Web Archive, n.d., FAQ).<sup>26</sup> These Regulations

---

<sup>24</sup> Theatres Act 1968, <https://www.legislation.gov.uk/ukpga/1968/54>

<sup>25</sup> Copyright Act 1911, <https://www.legislation.gov.uk/ukpga/Geo5/1-2/46/introduction/enacted>

<sup>26</sup> Legal Deposit Libraries Act 2003, <https://www.legislation.gov.uk/ukpga/2003/28/contents>



would not come into effect until 2013 through The Legal Deposit Libraries (Non-Print Works) Regulations 2013 (NPLD).<sup>27</sup> The 2003 Act also separated copyright law from legal deposit law (Muir, 2005, p. 3). Nonetheless, the 2003 Act did provide the legislative framework to enable the archiving of web content, albeit with permission from website owners (Bingham & Byrne, 2021, p. 2), and mandated the responsible minister with the powers to bring in regulations for digital collecting, including websites, under legal deposit which could be enacted at the appropriate time in the future.

Following a report commissioned by the Wellcome Trust and the Joint Information Systems Committee (JISC) (Day, 2003) on the feasibility of developing a UK web archiving service, six institutions came together to form UKWAC “to experiment with collection of website materials before the implementation of legal deposit legislation covering web publishing” (UK Web Archive, 2009, F.A.Q.); Bingham & Byrne, 2021, p. 2). The institutions included: The National Archives, the British Library, the national libraries of Scotland and Wales, the Wellcome Library and JISC (Bailey & Thompson, 2006). The UKWAC web archive was publicly launched in May 2005 with some of their earliest collections being the Indian Ocean Tsunami December 2004, the 2005 General Election and the 2005 July London terrorist attacks (Bailey & Thompson, 2006; UKWAC, 2005; UK Web Archive, 2020).<sup>28</sup>

It should also be noted however, that publishing, and communications technologies had rapidly advanced between the time the 2003 Act came into force in early 2004 and the time the 2013 NPLD Regulations came into effect (Arnold-Stratford & Ovenden, 2020, p. 5). From 2003 – 2013 the practical details for the NPLD regulations were worked out, during which time the UK Legal Deposit Libraries (LDLs) selectively archived websites under existing copyright law while contributing to the discussion on whether digital collecting needed legislation or whether it could be carried out under voluntary deposit. The conclusion was that seeking permission to archive from website publishers was not feasible and the regulations were necessary. The legislation was therefore updated in 2013 to allow for an annual web crawl of the UK web estate, including NI, undertaken by the UK Web Archive, a partnership of the six UK LDLs. The NPLD Regulations also solidified the establishment of the UK Web Archive (Bingham & Byrne, 2021, p. 2).

---

<sup>27</sup> The Legal Deposit Libraries (Non-Print Works) Regulations 2013, <https://www.legislation.gov.uk/uksi/2013/777/contents/made>

<sup>28</sup> With the passing of the 2013 NPLD Regulations, the UKWAC collections automatically transferred to the UK Web Archive under the partnership of the six UK legal deposit libraries.

In terms of collection, the UK Web Archive offers the following summary regarding their legal deposit collection strategy

As per the Non-Print Legal Deposit regulations we the six UK Legal Deposit Libraries are empowered to collect any and all UK based websites. In effect this includes all websites that have a UK top level domain name such as .UK, .SCOT, .WALES, .CYMRU and .LONDON plus any websites that are identified as being hosted on a server located physically in the UK via a geo-ip lookup. Additionally, if a website contains a UK postal address or the website owner confirms UK residence or place of business their website can be included. In order to build comprehensive thematic website collections, we occasionally request permission to archive non-UK websites from the site owner (UK Web Archive, n.d., Frequently Asked Questions).

Thus, the extent of the scope of their legal deposit collection efforts goes well beyond the capture of a national web domain demarcated by a basic ccTLD. Nonetheless, there are still “significant gaps in the heritage acquired as websites on ‘non-UK’ top level domain names, such as .com, are not automatically identified” (Bingham & Byrne, 2021, p. 2). This highlights the challenges for demarcating the geographical, structural and “imaginary” boundaries of a national web domain (Ben-David, 2019, pp. 89–91; Kahn, 2019, pp. 164–165; Webster, 2019, pp. 110–112).

In addition, the PRONI Web Archive commenced a selective web archiving initiative in 2010 to capture and preserve websites of NI government departments, local councils, public sector organisations and websites which have social, cultural, political, religious, or economic significance for the preservation of NI heritage. However, prior to 2013, the UK/NI web space was not systematically captured as part of legal deposit, and therefore much of the earlier NI webspace will have disappeared or changed drastically (Jackson, 2015a). To salvage some of the UK web estate prior to 2013, the Joint Information Systems Committee (JISC) acquired a dataset from the Internet Archive which included all .uk websites in their web archive collections that were crawled from 1996-2013 (UK Web Archive, n.d., JISC UK Web Domain Dataset).

The UK Web Archive is managed by the British Library, inclusive of its technical infrastructure (Pennock, 2013, p. 27). Initially, UKWAC utilised PANDAS software which was developed by the National Library of Australia, and the collections were “hosted by an external agency” (Pennock, 2013, p. 27). In 2008, UKWAC moved to an in-house operation using the Web Curator Tool (WCT) workflow management tool which was collaboratively developed by the British Library and the National Library of New Zealand through an International Internet Preservation Consortium (IIPC) funded project (Pennock, 2013, p. 27; Web Curator Tool, n.d.,

History). However, with the implementation of NPLD regulations in 2013, this brought about a major transformation in web archiving for the UK legal deposit libraries, “necessitating new workflows to deal with the selection, annotation and curation of content harvested both as part of the broad domain crawls and as part of more frequent and targeted crawling activity” (UK Web Archive, n.d., W3ACT User Guide). The WCT was not scalable, so the Annotation Curation Tool (W3ACT) was developed “to meet the requirements of subject specialists wishing to curate web content harvested under the Legislation” (UK Web Archive, n.d., W3ACT User Guide, also Jackson, 2016a). For example, W3ACT allows “users to perform numerous curatorial tasks including the assignation of metadata and crawl schedules to web content, quality assurance and the ability to request permission for open access to selected websites” (UK Web Archive, n.d., W3ACT User Guide). Additionally, the change in curation tool, also necessitated a change in crawling software from HTTrack to Heritrix (Pennock, 2013, p. 27).

The UK Web Archive are keen users of open-source tools and in doing so, they contribute back to the web archiving community (Jackson, 2022a; UK Web Archive, 2018). They also offer research support to PhD students wishing to use their collections. They have also collaborated with academic communities to develop tools for users such as SHINE. SHINE is a prototype of a potential research tool that can be used to access and analyse web archive data. It was developed as part of the Big UK Domain Data for the Arts and Humanities project funded by the UK Arts and Humanities Research Council. The data that underpins this service was acquired by JISC from the Internet Archive and includes all .uk websites in the Internet Archive web collection crawled from 1996 to April 2013, when NPLD came into effect (UK Web Archive, n.d., JISC UK Web Domain Dataset). The JISC UK Web Domain Dataset is available for use through the UK Web Archive website and listed in the British Library Shared Research Repository. Users can also search on SHINE either using a URL or keywords. The search results can then be further filtered by using predefined facets. In addition, trend graphs can be generated by using keywords and a time range between 1996 and 2013. By clicking on a single point in a trend graph this will generate a sample of 100 resources that reference that keyword and link out to the Internet Archive (UK Web Archive, n.d., SHINE; Byrne, 2019).<sup>29</sup> The JISC UK Web Domain Dataset (1996-2013) also offers four derived datasets which can be reused by researchers for big data analysis (UK Web Archive, n.d., JISC UK Web Domain Dataset). Furthermore, the dataset offers an example of how ‘some’ UK digital heritage was

---

<sup>29</sup> UK Web Archive SHINE, <https://www.webarchive.org.uk/shine>

salvaged for the years before the legislation allowed for selective web archiving from 2004, and for the crawling of the UK national web domain from 2013.

The UK Web Archive offers a selection of collections, with content both open access and onsite access, which may be representative of the NI web space. A few collections to note here are the UK General Election collections which cover the NI web space, and the Gender Equality collection which has a bodily autonomy subsection that covered the 8th Amendment Referendum and the Now for NI campaign. NI is further represented in sporting collections although there are some gaps. The News collection has a wide variety of NI publications. The Easter Rising 1916/2016 collection is also relevant for both NI and ROI. The UK Web Archive and PRONI have also collaborated on several curated web archive collections to offer a NI perspective on topics and events including Brexit, the 2019 UK General Election, Covid-19, UEFA Women's Euro England 2022, and these are available through the UK Web Archive (Murchan, 2020b). All published collections are visible through the UK Web Archive website on the Topics and Themes page (see [Figure 3.4](#)).<sup>30</sup>

End users should be aware that the records in the collections contain a mixed model in terms of access; some archived websites being open access and some only available onsite at legal deposit libraries. Nonetheless, it is useful to have collections arranged by topic and theme. This structure enables researchers to assess whether or not the holdings might warrant a trip to the TCD library to view the websites onsite at a library terminal. However, as mentioned previously in section 2.3, onsite access to the UK national web domain collections presents multiple challenges often due to the restrictive nature of the UK legal deposit legislation (NPLD). Indeed, Gooding et al. (2021) report that “many researchers have publicly questioned whether the restrictive access protocols for NPLD are in fact a barrier to the usage of electronic publications” (p. 1155). Furthermore, they suggest that the NPLD protocols are not in line with current trends in digital user expectations and information seeking behaviours (Gooding et al., 2019, p. 21). It should also be noted that by the time the 2013 NPLD Regulations had come into effect, publishing and communications technologies had rapidly advanced in the meantime (Arnold-Stratford & Ovenden, 2020, p. 5).<sup>31</sup>

---

<sup>30</sup> UKWA Topics and Themes, <https://www.webarchive.org.uk/en/ukwa/category/>

<sup>31</sup> Legal Deposit Libraries Act 2003 (Commencement) Order 2004, <https://www.legislation.gov.uk/uksi/2004/130/introduction/made>

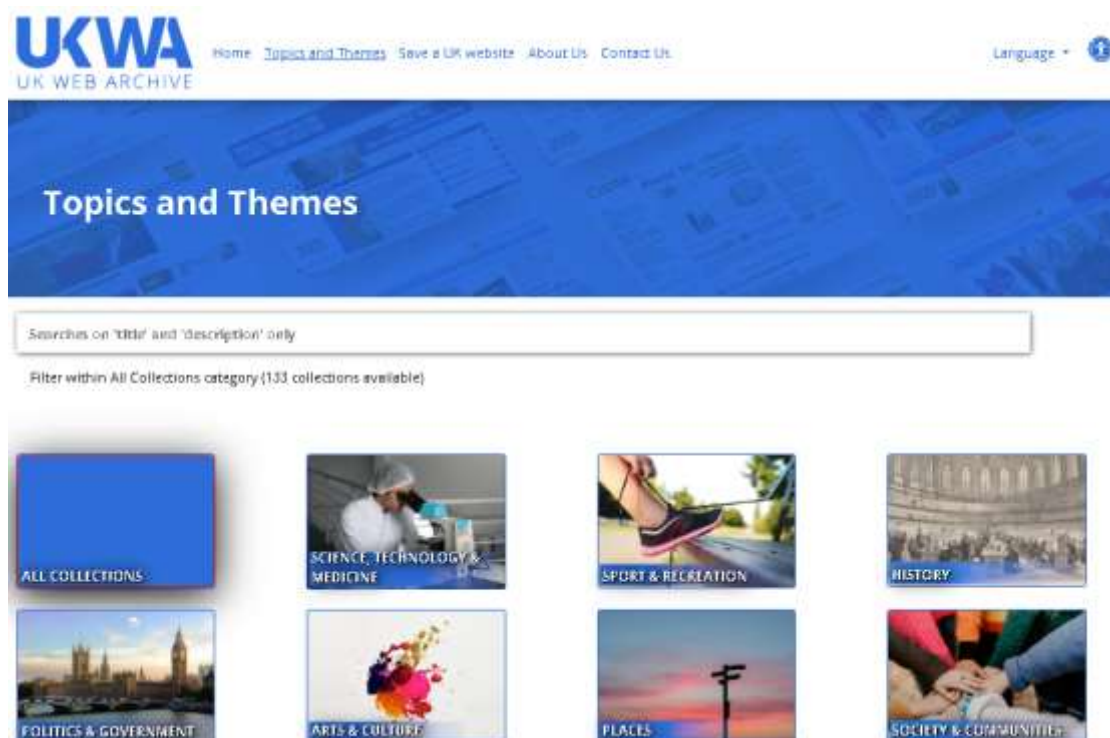


Figure 3.4: Screenshot of UKWA web page for Topics and Themes, taken on 2022-10-16

### 3.3.3 NI web space in brief

While there are resource and legislative limitations, concrete efforts are being made to provide a balanced approach towards the collection and preservation of the NI web space. Firstly, the UK Web Archive captures and preserves websites from the NI web space, through a selective collecting approach and through an annual domain crawl of the UK web space. This content is accessible onsite in UK legal deposit libraries, inclusive of the Library of TCD in Dublin. As outlined previously, only a small percentage of this collection has been made open access with consent from the content owners. Secondly, through a two-fold approach by PRONI to provide a publicly accessible selective web archive collection, through (i) the collection of websites of government, public bodies etc., with notifications of the intent to collect, and provisions of a takedown policy, and (ii) a permissions-based approach for privately funded websites. And thirdly, through a collaborative effort by the PRONI Web Archive and the UK Web archive for the development of curated collections. Access to the collections differ, with PRONI Web Archive being open access, and the UK Web Archive being a mix of both open and onsite access. However, as discussed earlier, onsite access presents multiple challenges for end users due to the restrictive nature of current UK legal deposit legislation, while onsite access is economically non-viable for many users.

### 3.4 ROI Web Space

In terms of a legal deposit strategy, it would make sense for the ROI to organise a collection development strategy, similar to the UK/NI model, for the capture and preservation of the ROI web space. In this way future generations will have access to a more representative landscape of Irish digital heritage on the web. However, there would need to be a more realistic approach towards the provision of access, considering the challenges that are currently being presented due to the restrictive nature of UK legal deposit legislation, which are in essence barriers for innovation. The JISC UK Web Domain Dataset (1996-2013) also offers a model for the salvage of ‘some’ digital heritage for the ROI’s .ie domain from 1996 to the current day.

In addition, while the UK model on demarcating a national web domain would be an equally good example for demarcating the Irish web domain, there is also the need to consider that there are “significant gaps in the heritage acquired as websites on ‘non-UK’ top level domain names, such as .com, are not automatically identified” (Bingham & Byrne, 2021, p. 2). To overcome this, it would be useful to also look at the Icelandic model which “contains all web sites hosted on the Icelandic domain .is” as well as many websites “hosted elsewhere that are in Icelandic or refer directly to matters of interest to Iceland” (IIPC, n.d., Landsbókasafn Íslands). The Icelandic Web Archive is also open access except for websites where the user must pay for access and when the content owners have requested that access to the archived version of the content is blocked (IIPC, n.d., Landsbókasafn Íslands).

Thus, the ROI should evaluate the demarcation of their national web domain to ensure such gaps are minimised. However, at the current time, the inclusion of a national web domain archive is not part of ROI legal deposit legislation, nor does there seem to be any efforts (in the public domain at least) towards the salvage of ROI digital heritage from any other web archive. Moreover, for the most part, the parliamentary discussions regarding the establishment of a national web domain archive are solely focused on the ROI’s .ie domain ([Figure 3.10](#)).

While many western societies undertook a review of their copyright, heritage, and legal deposit laws from the 1990s, due to advances in publishing and communication, and many have implemented reform to account for legal deposit of non-print materials, including the development of national web archiving programmes through a selective based approach, and routine national domain web archiving, the ROI has trailed behind Canada, New Zealand, and much of Europe (Conul, 2012, p. 14).

### 3.4.1 NLI web archive

As mentioned before, the NLI is the official library of record in the ROI and is responsible for collecting for, and on behalf of the Irish state (NLI, Collection Development Policy 2022-2026). The NLI commenced a Born Digital Programme in 2011 to

identify the role of the National Library of Ireland in relation to the collection of born-digital material and to identify, collect and make accessible born-digital material as part of day-to-day collection development activities (NLI Annual Report 2011, p. 8).

As part of this, following public procurement processes the NLI teamed up with the IMF subscription-based web archiving service in Europe for a selective (permissions-based) web archiving initiative for the 2011 General Election. The project entailed the capturing of 100 websites to include: candidate's sites, political blogs, news media, and some official government sites. Later the same year, web content was collected for the 2011 Irish Presidential Election (NLI Annual Report, 2011, p. 8). Since then, the NLI has continued to operate a small-scale selective web archiving programme for the capture of Irish social, cultural, political, governmental, and commemorative heritage on the web. The collections are openly accessible via the NLI Web Archive. Following a procurement process in early 2018, the NLI then commenced using the Internet Archive's Archive-It subscription service and migrated their holdings from the IMF platform to its current location on the Archive-It platform. Figure 3.5 offers a screenshot of the initial public NLI Web Archive via the IMF interface from 2015. Also worth mentioning is the fact that the migration of the NLI Web Archive from one platform to another would have caused some disruption for end users who were using the initial IMF resource for research at the time (Aturban, 2019a). Ultimately all the URLs linking to the data in the original NLI Web Archive through the IMF platform became invalid (<http://collection.europarchive.org/nli/>), while the Archive-IT web archive offers a new URL (<https://archive-it.org/home/nli>).

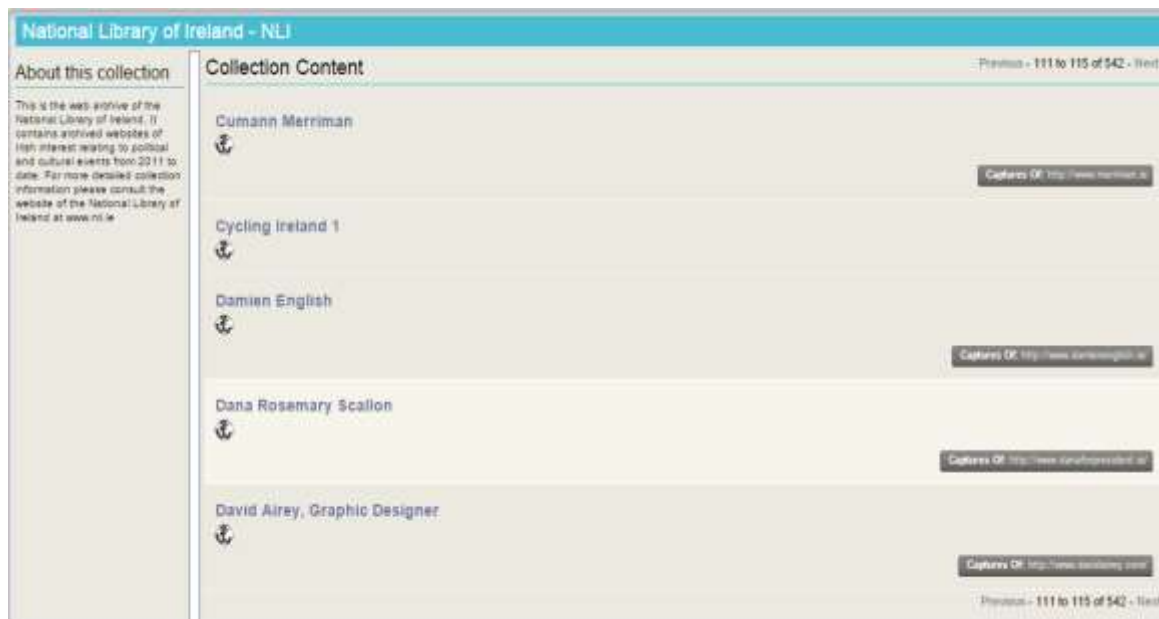


Figure 3.5: Screenshot of NLI Web Archive user interface on the Internet Memory Foundation platform, taken in June 2015 (personal archive)

In January 2023, the selective NLI Web Archive contained 75 collections with 3,114 websites overall which cover a diverse range of topics such as Irish History, LGBTI+, Irish Entertainment, Ageing in Ireland, Coastal & Island Life, Irish Music, Higher Education, Agriculture, Horticulture & Food, and Housing & Property, to name but a few. One can browse the collections alphabetically or browse an alphabetical listing of the individual websites (Figure 3.6 & Figure 3.7). Both the pre-2018 archived websites (migrated from the IMF platform) and the post-2018 archived websites are organised by collection on the Archive-It platform with applicable metadata which is very beneficial for end users (see for example the General Election 2011 collection).<sup>32</sup> Indeed, the application of metadata is a noteworthy undertaking by the NLI considering it is a resource intensive undertaking (Costa, 2021, p. 72), coupled with the fact that the NLI is already “constrained primarily by limited staff numbers” (Collins, 2018, p. 180). Equally useful for understanding some of these earlier collections are the NLI selective web collection development policies (NLI, 2022, Selective Web Archive Collections). The resource can also be searched by inputting a known URL, and through a full-text search.

<sup>32</sup> NLI, Archive-It - General Election 2011, <https://archive-it.org/collections/19959>





Figure 3.6: Screenshot of NLI Web Archive interface on the Archive-It platform showing a total of 3,114 websites in their collections, taken on 2023-01-16

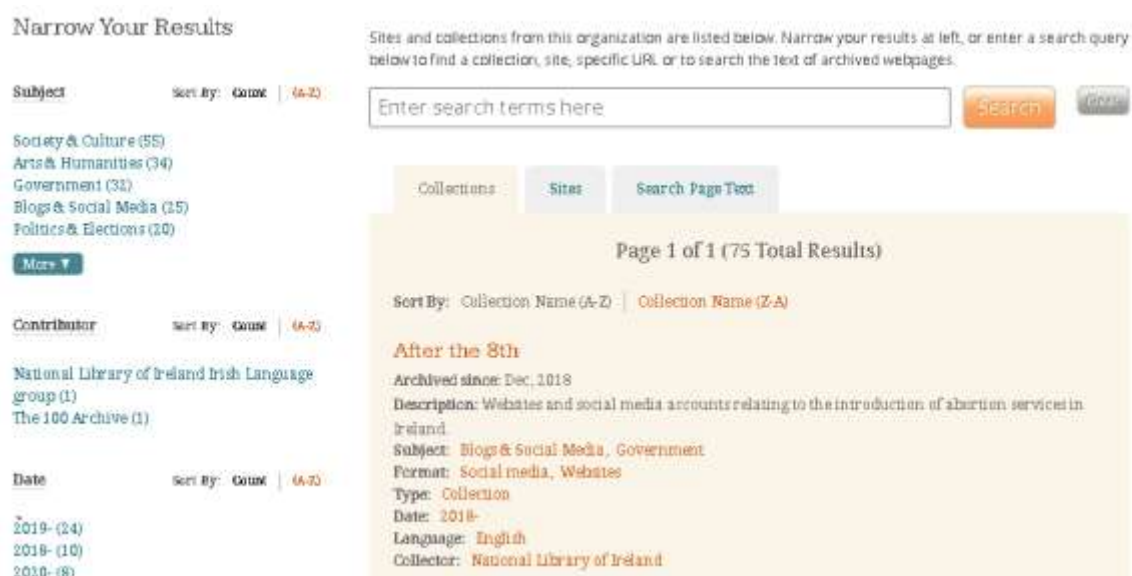


Figure 3.7: Screenshot of NLI Web Archive interface on the Archive-It platform showing a total of 75 collections, taken on 2023-01-16

Further to this, following public procurement processes the NLI collaborated with the Internet Archive web archiving service to capture the Irish ccTLD web domain (.ie) in 2007 and again in 2017 (NLI, 2017; NLI, n.d., Irish Domain Web Archive). In 2017, *The Irish Times* online, speculated that the domain crawls would become available for access onsite in the NLI reading rooms (Taylor, 2017a). However, this has yet to happen, and there have been no further domain crawls conducted since 2017. The issue here for both the archiving of the Irish

web domain (.ie), and the provision of access to the domain content that is already captured, is due to the current state of Irish legislation on copyright and legal deposit.

While the NLI is a legal deposit library, digital legal deposit legislation was not enacted in Ireland at the time the 2007 or 2017 domain crawls were conducted (O'Dell, 2018; Collins, Joint Committee on Tourism, Culture, Arts, Sport and Media, 13 October 2021). Moreover, while digital legal deposit legislation came into force in December 2019 through the Copyright and Other Intellectual Property Law Provisions Act 2019 (hereafter, COIPLPA, 2019), it did not include a clause for crawling the Irish national web domain. Thus, the 2007 and 2017 domain crawls remain inaccessible to both researchers and the public alike, and the NLI is prohibited from doing any further domain crawls. However, COIPLPA (2019) does contain a clause as follows:

Within twelve months of the enactment of this Act the Government shall bring forward a report on the feasibility of establishing a digital legal deposit scheme to serve as a web archive for .ie domain contents and advise on steps taken towards that goal (Section 108: COIPLPA, 2019; [Figure 3.8](#)).

Regrettably, to date, the ROI government has failed to deliver a feasibility report and thus failed to incorporate a domain web archive as part of copyright and legal deposit legislation. Consequently, the NLI continues to be prohibited from doing any further crawls, while the country continues to suffer from mass losses of Irish digital heritage. For more understanding of these issues, we offer an overview of copyright and legal deposit legislation in the ROI.



Figure 3.8: Screenshot of Section 108 in the Copyright and Other Intellectual Property Law Provisions Act 2019, taken on 2022-10-07

Following partition, a legal deposit scheme for print publications was introduced in 1927 through the Industrial and Commercial Property (Protection) Act, 1927 (ICPPA, 1927).<sup>33</sup> This established the relationship between copyright and legal deposit in the same legislation. The ICPPA (1927) also formalised the NLI as a legal deposit library alongside the Library of TCD, the British Museum (now the British Library), and the three constituent college libraries of the National University of Ireland (at the time), being University College, Dublin, University College, Cork, and University College, Galway (Section 178, Ireland, ICPPA, 1927). The next change in the law which applied to legal deposit was in Section 56 of the Copyright Act, 1963, whereby it included the library of St. Patrick's College, Maynooth (later Maynooth University) as a designated legal deposit library.<sup>34</sup> Further amendments to legal deposit legislation saw the introduction of two new legal deposit libraries through Section 7 of the University of Limerick Act, 1989, and Section 6 of the Dublin City University Act, 1989.<sup>35</sup> Currently, the ROI legal deposit institutions include: the National Library of Ireland, the Library of Trinity College Dublin, the British Library, and the libraries of Dublin City University, National University of Ireland, Galway, Maynooth University, University College Dublin, University College Cork, and University of Limerick. In addition, Irish publishers may also be obliged to deposit publications, on request by the Bodleian Library, University of Oxford, the Cambridge University Library, the National Library of Scotland, and the National Library of Wales (Section 198: Copyright and Related Rights Act, 2000; National Library of Ireland, 2014).<sup>36</sup>

While there were subsequent amendments to ROI copyright law during the 1990s, the amendments did not affect legal deposit per se. Similar to the UK, Ireland was a member of the EEC since 1973 and was obliged to keep in line with their directives. For example, the European Communities (Legal Protection of Computer Programs) Regulations, 1993 was introduced by the ROI government to accommodate copyright for computer programmes and

---

<sup>33</sup> Industrial and Commercial Property (Protection) Act, 1927,  
[https://www.bailii.org/ie/legis/num\\_act/1927/0016.html#zza16y1927](https://www.bailii.org/ie/legis/num_act/1927/0016.html#zza16y1927)

<sup>34</sup> Copyright Act, 1963, <https://www.irishstatutebook.ie/eli/1963/act/10/enacted/en/html>

<sup>35</sup> University of Limerick Act, 1989,  
<https://www.irishstatutebook.ie/eli/1989/act/14/enacted/en/html>;  
 Dublin City University Act, 1989, <https://www.irishstatutebook.ie/eli/1989/act/15/enacted/en/html>

<sup>36</sup> Copyright and Related Rights Act, 2000, Section 198,  
<https://www.irishstatutebook.ie/eli/2000/act/28/section/198/enacted/en/html>

programmers in compliance with EEC Council Directive 1991/250/EEC.<sup>37</sup> And like the UK, the ROI also had to comply with the EEC Council Directive No. 93/98/EEC, to harmonise copyright across member states, and thus, the Irish government introduced the European Communities (Term of Protection of Copyright) Regulations, 1995.<sup>38</sup>

The Copyright and Related Rights Act, 2000 (CRRA) was also established in some parts to comply with EC directive 2001/29/EC – Harmonisation of certain aspects of copyright and related rights in the information society.<sup>39</sup> Furthermore, section 199 of the CRRA recognises the need for the inclusion of non-print materials such as:

any engraving, photograph, text of a play, cinematograph film, microfilm, video recording, sound recording, record, diskette, magnetic tape, compact disc, or other thing on or in which works or information or the representations thereof is written, recorded, stored or reproduced but does not include local records or local archives within the meaning, in each case, of section 65 of the Local Government Act, 1994, or books within the meaning of section 198 of the Copyright and Related Rights Act, 2000 (Section 199: CRRA, 2000).

Whereas section 198 of the CRRA relates to the legal deposit of books, defining a book to include:

every part or division of a book, pamphlet, sheet of letterpress, sheet of music, map, plan, chart or table separately published, but shall not include any second or subsequent edition of a book unless such edition contains additions or alterations either in the letterpress or in the maps, plans, prints or other engravings belonging thereto (Section 198, CRRA, 2000).

The CRRA also makes an acknowledgement for the allowance of the deposit of books in electronic form in Section 198, as outlined below, however, the CRRA did not consider the

---

<sup>37</sup> European Communities (Term of Protection of Copyright) Regulations, 1995, <https://www.irishstatutebook.ie/eli/1995/si/158/made/en/print>; EEC Council Directive 93/98/EEC, <http://data.europa.eu/eli/dir/1993/98/oj/eng>

<sup>38</sup> European Communities (Term of Protection of Copyright) Regulations, 1995, <https://www.irishstatutebook.ie/eli/1995/si/158/made/en/print>; EEC Council Directive 93/98/EEC, <http://data.europa.eu/eli/dir/1993/98/oj/eng>

<sup>39</sup> Copyright and Related Rights Act, 2000, <https://www.irishstatutebook.ie/eli/2000/act/28/enacted/en/html>;

EC Council Directive 2001/29 - Harmonisation of certain aspects of copyright and related rights in the information society, <http://data.europa.eu/eli/dir/2001/29/oj/eng>

fact that some publications are born digital only, with no print counterpart, such as internet journals and publications, e-zines, or web pages.

(11) Where a copy of a book requested under subsection (1) is delivered in a form other than an electronic form, the Board or other authorities referred to in subsection (1) may request, in addition to that copy, a copy in an electronic form readable by means of an electronic retrieval system and on such request being made a copy in electronic form shall be delivered by the publisher to the Board or authority concerned.

(12) For the purposes of this section, “publication”, in relation to a book—

- a) means the issue of copies to the public, and
- b) includes its making available to the public by means of an electronic retrieval system,

and related expressions shall be construed accordingly (Section 198: CRR, 2000).

Recognising the need for measures to preserve Irish digital heritage, both the Library of TCD and the NLI, as nominated legal deposit libraries in the ROI, instigated different schemes to accommodate the collection of electronic and born digital publications. In the NLI Collection Development Policy 2009-2014, the NLI highlights: “If the challenge of collecting Ireland’s online presence is not addressed, it can be argued that the Library is not respecting its remit as a national memory institution” (p. 9). In a bid to provide some solutions, the NLI set up the Born Digital Programme in 2011, for the purpose of identifying and collecting born digital content and implementing web archiving practices as a regular activity of the library (NLI Annual Report 2011, p. 8; NLI Annual Report 2012, p. 13). The Library of TCD set up a voluntary electronic deposit scheme in Ireland through their resource [edepositIreland](#), as a “self-deposit service [...] open to all publishers in Ireland, be they individuals, local groups, publishing houses or organisations, who wish to share their publications with the world” (Library of TCD, n.d., [edepositIreland](#)). Also, as mentioned earlier the Library of TCD provides access to the UK Web Archive legal deposit domain web archive, which is accessible onsite through the library pc terminals.

It was not until 2011 that we witnessed a more serious investment by the ROI government (a coalition of the Fine Gael and Labour Party) to address both copyright and legal deposit due to advances in the web and the internet (O’Dell, 2013). In May 2011, Richard Bruton (Fine Gael), who was the Minister for Jobs, Enterprise, and Innovation at the time, established the Copyright Review Committee (CRC) to:

Examine the present national copyright legislation and identify any areas that are perceived to create barriers to innovation [and] Identify solutions for removing these barriers and make recommendations as to how these solutions might be implemented through changes to national legislation (CRC, 2012, p. 1; Department of Jobs, Enterprise, and Innovation, 2012, Consultation).

The CRC held a public meeting in July 2011 in Dublin and from there solicited more than 100 written submissions regarding copyright review overall. The submissions were at one time available to view on the website for the Department of Jobs, Enterprise and Innovation, but the website is no longer available on the live web, as the department was reformulated twice since then, which unsurprisingly meant new URLs for their websites (see [Table 3.1](#)). Therefore, one needs to use the NLI Web Archive to view the different types of stakeholders who made submissions.<sup>40</sup> It is also worth discussing the issues with ROI government department websites a little further.

Table 3.1 shows how one ROI government department was reformulated five times from 1997 to 2020, resulting in five different titled websites, of which the latest website only lists news/media items as far back as 2016, meaning that news/media items from 1997-2015 may only be found in a web archive, if at all. While the NLI has also been capturing ROI government department websites since 2011, scholars will need to rely heavily on the Wayback Machine for Irish web content created from the mid-1990s to 2011, in a bid to find fragments of what was at one time public information provided by ROI government departments on their websites ([Figure 3.9](#)). However, the Wayback Machine may only hold surface pages of an individual departmental website, and not have crawled the multitude of internal hyperlinks for departmental news or publication items for example. Moreover, there are undoubtedly a multitude of academic publications and government publications which contain URL references, linking to these departmental websites over the years which are no longer valid, and as Spinellis (2003) points out “references that cannot be located seriously undermine the foundations of modern scientific discourse” (p. 71).

---

<sup>40</sup> Department of Jobs, Enterprise and Innovation, Submissions Received by the Copyright Review Committee, NLI Web Archive, 2012, [https://wayback.archive-it.org/org-1444/20120613230622/http://www.djei.ie/science/ipr/crc\\_submissions2.htm](https://wayback.archive-it.org/org-1444/20120613230622/http://www.djei.ie/science/ipr/crc_submissions2.htm)



Department of Enterprise,  
Trade and Employment  
Timestamp: 1999-10-04  
([www.entemp.ie](http://www.entemp.ie))



Department of Enterprise,  
Trade and Employment  
Timestamp: 2002-05-29  
([www.entemp.ie](http://www.entemp.ie))



Department of Enterprise,  
Trade and Innovation  
Timestamp: 2010-05-07  
([www.deti.ie](http://www.deti.ie))



Department of Jobs,  
Enterprise, and Innovation  
Timestamp: 2011-06-06  
([www.djei.ie](http://www.djei.ie))



Department of Jobs, Enterprise  
and Innovation  
Timestamp: 2012-06-10  
([www.enterprise.gov.ie](http://www.enterprise.gov.ie))



Department of Business,  
Enterprise and Innovation  
Timestamp: 2017-09-05  
([www.dbei.gov.ie](http://www.dbei.gov.ie))

Figure 3.9: The changing nature of ROI Government Department websites and URLs is revealed by examining the captures of their websites in the Wayback Machine (see Table 3.1)

Table 3.1: Renaming of Department of Jobs, Enterprise and Innovation before and after formulation (Sources: Wikipedia, 2005+)

Dates	Department name	Website URL	URL Status notes	Link Rot
July 1997	Renamed as the Department of Enterprise, Trade and Employment	<a href="http://entemp.ie">http://entemp.ie</a>	Redirects to <a href="https://www.iva-advice.co">https://www.iva-advice.co</a>	YES
May 2010	Renamed as the Department of Enterprise, Trade and Innovation	<a href="http://www.deti.ie">http://www.deti.ie</a>	Page not Found	YES
June 2011	Renamed as the Department of Jobs, Enterprise and Innovation	<a href="http://www.djei.ie">http://www.djei.ie</a>	Redirects to new website, <a href="https://enterprise.gov.ie/djei/en/">https://enterprise.gov.ie/djei/en/</a>	YES

Sept. 2017	Renamed as the Department of Business, Enterprise and Innovation	<a href="https://dbei.gov.ie/en">https://dbei.gov.ie/en</a>	Redirects to new website, <a href="https://enterprise.gov.ie/djei/en/">https://enterprise.gov.ie/djei/en/</a>	YES
Nov. 2020	Renamed as the Department of Enterprise, Trade and Employment	<a href="https://enterprise.gov.ie/djei/en/">https://enterprise.gov.ie/djei/en/</a>	Still there, pending migration to the central government website	

Getting back to the CRC and copyright review, from an analysis of the written submissions on copyright review, the CRC published a consultation paper in 2012 which proposed amendments to copyright legislation and solicited further consultations and feedback for the proposals (CRC, 2012, Copyright and Innovation: A Consultation Paper). Of interest in the consultation paper is the classifications of the submissions into the categories of “(i) rights-holders; (ii) collecting societies; (iii) intermediaries; (iv) users; (v) entrepreneurs; and (vi) heritage institutions” (O’Dell, 2012; CRC, 2012, pp. 9–10). The Consultation Paper breaks these categories as outlined verbatim below, which offers us an overview of the stakeholders with an interest in Irish copyright review.

- rights-holders: this category includes the people who create the copyright work, and as well as their publishers, music labels, movie studios, broadcasters and so on,
- collecting societies: this category includes societies which grant licences of copyrighted works and collect copyright royalties for distribution back to the rights-holders,
- intermediaries: this category includes internet service providers, online search engines, social networks, and trading sites,
- users: this category includes the consumers, purchasers and users of copyright works,
- entrepreneurs: this category includes online start-ups,
- heritage institutions: this category includes libraries, archives, galleries, museums, schools, universities and other educational establishments, and the like (Copyright and Innovation: A Consultation Paper, pp. 9–10).

Following further analysis, the CRC then published the Modernising Copyright report in October 2013, which offered modern solutions to Ireland’s outdated copyright laws.

Of particular interest, the Modernising Copyright report makes recommendations for the introduction of digital legal deposit to current legal deposit institutions, and further to this, that such institutions should be permitted to “make copies of our online digital heritage by



reproducing any work that is made available in the State through the internet” (CRC, 2013, p. 5). In this context, the report clarifies the meaning of a work on the internet to be:

a work shall have been made available in the State through the internet where (a) it is made available to the public either from a website with a domain name which relates to the State or to a place within the State, or by similar or related means, or (b) it is made available to the public either by a person any of whose activities relating to the creation or the publication of the digital publication takes place within the State, or by a person with similar or related connections to the State (CRC, 2013, p. 153).

The report further advocates for the “formation of a Copyright Council of Ireland, as an independent self-funding organisation, created by the Irish copyright community, recognised by the Minister, and supported and underpinned by clear legislative structures provided” (CRC, 2013, p. 9). The purpose of which would serve to

ensure the protection of copyright and the general public interest as well as encouraging innovation; and it should have a broad subscribing membership and a Board drawn widely from the Irish copyright community. It should provide education and advice on copyright issues, advocate both nationally and internationally for developments in copyright policies or procedures, and work towards solutions on difficult copyright issues. It should be able to establish a Digital Copyright Exchange (to expand and simplify the collective administration of 10 copyrights and licences), a voluntary alternative dispute resolution service (to meet the need for an expeditious dispute resolution service outside the court system), and an Irish Orphan Works Licensing Agency (to provide a solution to the problem of orphan works) (CRC, 2013, p. 9).

Indeed, a Copyright Council of Ireland would also make sense in terms of the need for a continual evaluation of legal deposit legislation in line with the fragility of born digital heritage and the technological advances in publishing and communication technologies. Nonetheless, the Irish government was slow to embrace the zeitgeist of the recommendations, despite real-time concerns for the loss of Irish digital heritage on the web in the meantime.

Copyright amendments were imminent, due to the advances in the web and the internet in Ireland from early 2000 (Sterne, 2015+), and the prevalence of copyright infringements online from mid-2000 onwards (e.g., music, film, photography, etc.) (O’Dell, 2013; Morris, 2019). There was also a need for changes to copyright to facilitate disability requirements and educational needs, such as allowing for the use of copyrighted multimedia on a classroom whiteboard, and for allowing the modification of books to meet the needs of individuals with disabilities in line with the international Marrakesh Treaty to Facilitate Access to Published

Works for Persons who are Blind, Visually Impaired, or otherwise Print Disabled (Department of Enterprise, Trade and Employment, 2016). However, the inclusion of “digital” with regards to legal deposit could not be assumed. As O’Dell (2016) points out, when the government finally announced the drafting of the Copyright and Related Rights (Miscellaneous Provisions) Bill, 2016, they were initially opting for the incorporation of a digital legal deposit scheme “on a voluntary basis” (O’Dell, 2016; Department of Enterprise, Trade and Employment, General Scheme, 2016).

Yet another consultation was launched in April 2017, by the reformulated Department of Arts, Heritage, Regional, Rural and Gaeltacht Affairs (previously Department of Arts, Heritage and the Gaeltacht) in consultation with the NLI. Titled, ‘Consultation on the Legal Deposit of published digital material in the 21st century in the context of Copyright legislation’, it was aimed “at gathering stakeholder views in regard to whether or not the policy in relation to Legal Deposit should include the collecting, preserving and making available of all contemporary publication formats, including online digital formats such as websites.” The consultation requested opinions from “the library and archives community, publishers and members of the public in the context of the review of the Copyright and Related Rights Act” (Department of Arts, Heritage, Regional, Rural and Gaeltacht Affairs, 2017).

In August 2017, the Department of Arts, Heritage, Regional, Rural and Gaeltacht Affairs was reformulated to become the Department of Culture, Heritage, and the Gaeltacht, and would retain responsibility for addressing legal deposit. In December 2017, the reformulated department issued a response to the 42 submissions they received for the consultation process. Regarding the extension of legal deposit to include digital formats inclusive of websites, the Department responds with the following statement:

93% of responses to this question were strongly supportive, with respondents highlighting the ephemeral nature of online digital material, and the huge threat of its loss, unless institutions such as the National Library of Ireland operating in the cultural heritage area are legally mandated to preserve it. Many respondents also referred to the fact that the history of the 21st century is recorded online, and how the loss of this online information will lead to the loss of ‘significant national documentation’ as well as the loss to researchers of the outputs of research (Department of Culture, Heritage, and the Gaeltacht, 2017).

Thus, there were recommendations by the *Modernising Copyright* (2013) report for the establishment of an Irish web domain archive, and 93% of the submissions to the Consultation on Legal Deposit in 2017, were also in support of the establishment of an Irish web domain archive.

The following year saw the introduction of the Copyright and Other Intellectual Property Law Provisions Bill 2018 (as initiated) which was brought before Dáil Éireann in March 2018,

to amend the Copyright and Related Rights Act 2000 to take account of certain recommendations for amendments to that Act contained in the Report of the Copyright Review Committee entitled ‘Modernising Copyright’ published by that Committee in October 2013 and also to take account of certain exceptions to copyright permitted by Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society (Copyright and Other Intellectual Property Law Provisions Bill 2018 (as initiated), 2018).

Nonetheless, while the Bill (as initiated) “extended the copyright deposit regime to ebooks”, it did not “provide for the harvesting of the .ie domain” (O’Dell, 2018). Fianna Fáil then put forward the following amendment which was approved by Dáil Éireann.

Within twelve months of the enactment of this Bill the Government shall bring forward a report on the feasibility of establishing a digital legal deposit scheme to serve as a web archive for .ie domain contents and advise on steps taken towards that goal (amendment qtd. in O’Dell, 2018).

O’Dell suggests that this was at least progress, “even if it amounted to making haste slowly” (2018). However, at the Seanad Éireann committee stage, Senator Fintan Warfield (Sinn Féin) insisted that

The time for examining feasibility has long passed. It should have been done as soon as it was recommended [in 2013]. The only way we can have certainty in respect of this issue is to provide for it in law through this Bill and I respectfully encourage the Government to do so (Fintan Warfield, Seanad Éireann, Copyright and Other Intellectual Property Law Provisions Bill 2018: Committee Stage, 03 October 2018).

Thus, Warfield proposed an amendment as outlined below.

(c) For the purposes of this subsection, a work shall have been made available in the State through the internet where—

- (i) it is made available to the public either from a website with a domain name which relates to the State or to a place within the State, or by similar or related means, or
- (ii) it is made available to the public either by a person any of whose activities relating to the creation or the publication of the digital publication takes place within the State, or by a person with similar or related

connections to the State (Fintan Warfield, Seanad Éireann, Copyright and Other Intellectual Property Law Provisions Bill 2018: Committee Stage, 03 October 2018).

However, the Minister of State for Training, Skills, Innovation, Research and Development, John Halligan (Independent) had responsibility for carrying the Bill on behalf of the Fine Gael and Independent coalition government and objected to Warfield's amendment as outlined below.

Providing for a full digital deposit system that would facilitate capturing the web is not simply a matter of changing copyright legislation. It is a significant national project that requires multi-institutional collaboration, significant resources and Skillsnet [*sic*] for capturing and preserving Ireland's digital record, according to my advice. I reiterate that this is a matter for the Minister for Culture, Heritage and the Gaeltacht who has responsibility for policy in this area. My Department and the Department of Culture, Heritage and the Gaeltacht have actively worked together on that matter for some time and we will continue to do so until the robust regulatory framework is developed. We will facilitate the necessary corresponding legislation amendments in due course [...]

This work, however, is not yet sufficiently progressed for any technical amendments to copyright law. As that is the final aspect of the project, now that all the necessary due diligence has been done, Government mechanisms have been established and funding has been agreed with the Minister for Public Expenditure and Reform, it is not possible for amendments to copyright law to be progressed in isolation from Government approval for the project as a whole [...]

[A] new section stipulates that a report be published within 12 months of the Bill being enacted. This was accepted by all parties and viewed as a pragmatic way to advance the project while allowing time for the necessary work to take place in the Department of Culture, Heritage and the Gaeltacht, in co-operation with my Department and the Department of Public Expenditure and Reform. The House can rest assured that both Departments are actively engaged in advancing the proposal and the report will be prepared within 12 months, as specified in the Bill (John Halligan, Seanad Éireann, Copyright and Other Intellectual Property Law Provisions Bill 2018 – Committee Stage, 03 October 2018).

Nonetheless, Senator Warfield “pressed it to a vote” (O'Dell, 2018). According to O'Dell (2018): “On the electronic vote, there was a tie – Tá (yes) 18; Níl (no) 18 – and the amendment was defeated on the casting vote of the Leas Cathaoirleach (Deputy Speaker).” However, Warfield “called for a walk-through vote, and the amendment was [carried] – Tá (yes) 19; Níl

(no) 17” (O’Dell, 2018). Nonetheless, when it went to Report Stage in Dáil Éireann in May 2019, Minister Halligan

unapologetically restated his objections that there were issues with other government departments and public institutions, and that it would have significant resource implications, and he put down an amendment to reverse Senator Warfield’s earlier successful amendment (O’Dell, 2019).

Finally, a digital legal deposit scheme was formally organised in the ROI through the Copyright and Other Intellectual Property Law Provisions Act 2019 (COIPLPA, 2019).<sup>41</sup> The Act allows for the collection of e-books and journals on the internet, but it does not allow for the archiving of the Irish national web domain (Ryan et al., 2022). It does, however, provide a clause to “bring forward a report on the feasibility of establishing a digital legal deposit scheme to serve as a web archive for .ie domain contents and advise on steps taken towards that goal” within twelve months of the Act coming into force in December 2019. Yet, as of May 2023, a feasibility report has still not been produced.

Certainly, a feasibility study may have been disrupted due to the onset of the COVID-19 pandemic in Ireland in March 2020 (O’Dell, 2020). However, as the country gets back to normal, the pandemic should no longer be a reason for the holdup. ROI Parliamentary questions, and committee debates, also provide some indications as to the hold-up in adopting routine domain web archiving as a necessary component of a modern-day legal deposit scheme, and why the domain crawls already conducted by the NLI are inaccessible to researchers and members of the public.<sup>42</sup> Such reasons include the need to have consultation with multiple stakeholders, such as the publishing, heritage communities, and other government departments. There is also the question of the capacity of the NLI, and thus, a feasibility study should include details of the capacity of the NLI to take on a digital legal deposit web archive, in terms of infrastructure, technology, human resources, etc. For example, on 09 September 2021, in a Dáil Éireann session on parliamentary questions, TD Rose Conway-Walsh (Sinn Féin) put a question to the Minister for Tourism, Culture, Arts, Gaeltacht, Sport and Media, who at the time was (and still is at the time of writing) TD

---

<sup>41</sup> Copyright and Other Intellectual Property Law Provisions Act 2019, <https://www.irishstatutebook.ie/eli/2019/act/19/enacted/en/index.html>

<sup>42</sup> Examples here are added in the Bibliography and include: Dáil Éireann. (2021). Digital Archiving; Dáil Éireann. (2021). Intellectual Property; Joint Committee on Tourism, Culture, Arts, Sport and Media. (2021b). Key Priorities and Legislation of the Department of Tourism, Culture, Arts, Gaeltacht, Sport and Media; Seanad Éireann. (2021). An tOrd Gnó - Order of Business; Seanad Éireann. (2021). Nithe i dtosach suíonna - Commencement Matters – Digital Archiving.

Catherine Martin (Green Party), requesting when the Minister would produce the “report on the feasibility of establishing a digital legal deposit scheme for large-scale, systematic and sustained archiving of the Irish web domain” (Rose Conway-Walsh, Dáil Éireann, Digital Archiving, 09 September 2021). Minister Martin responds as follows:

My Department is working with the National Library of Ireland (NLI) on exploring the feasibility of expanding the NLI’s capacity to establish a digital legal deposit scheme to serve as a web archive for the .ie domain and work is ongoing. There are differing viewpoints on the introduction of digital legal deposit and it is important that consultations incorporate all viewpoints. It is hoped to bring forward a report in the coming months (Catherine Martin, Dáil Éireann, Digital Archiving, 09 September 2021).

Lyman (2002) also points to the need for “a process of negotiation among interested parties” when developing a web archive.

Furthermore, it will be equally important that discussions and negotiations take place with a diverse range of stakeholders. For example, representatives from the teaching and education sector, academics across a multitude of disciplines, and other types of end users such as public administrators, journalists, legal professionals, web designers, computer scientists, data analysts and local historians (Healy et al. 2022, p. 26; p. 102; p. 122; Ramesh & Hern, 2013; Winters, 2017; Truman, 2016, pp. 29–30; Bailey, 2015). It would also be beneficial to hold dialogues with information professionals from other national libraries who have experienced the transition from small-scale selective web archiving to large-scale web domain archiving and can advise on the challenges inclusive of how legislation impacts on implementation and use (Gooding et al., 2019). Most importantly, we need to value the opinions of information professionals with experience of working in Irish libraries and information ecosystems, and the Irish archives sectors. One example of this is evident from a sitting of the Joint Committee on Tourism, Culture, Arts, Sport and Media for a discussion on Engagement with Chairperson Designate of the Board of the National Library of Ireland (Joint Committee on Tourism, Culture, Arts, Sport and Media, 13 October 2021).

During the session, Mr. Eoin McVey, the NLI Chairperson (at the time), and Dr Sandra Collins, the NLI Director (at the time), were invited to discuss the challenges and achievements of the NLI (Joint Committee on Tourism, Culture, Arts, Sport and Media, 13 October 2021). As part of this, TD Johnny Mythen (Sinn Féin) asked the NLI representatives:

What are the legal obstacles in the way of archiving material through digital content? Does this need to be changed as soon as possible? What adequate

supports does the library need that are not in place now? (Johnny Mythen, Joint Committee on Tourism, Culture, Arts, Sport and Media, 13 October 2021).

Dr Collins's replies are worthy of reproducing below:

This is a critical issue for us. We collect one copy of every book published in the State, through copyright legislation legal deposit. We need to acknowledge the importance of content published on websites. Websites are a record of Irish life and we need to be able to make a copy of them and store and preserve them for future use and access. Section 108 of the Copyright and Other Intellectual Property Law Provisions Act 2019 is important. It allows for a report to be brought to Cabinet on the feasibility of a digital web archive. We are working with our parent Department, the Department of Tourism, Culture, Arts, Gaeltacht, Sport and Media, to bring that report to Cabinet. It is critical to us (Collins, Joint Committee on Tourism, Culture, Arts, Sport and Media, 13 October 2021).

Dr Collins further expands on these issues claiming that each year, "50% of Irish websites vanish forever or are changed so that they are unrecognisable from what they are now. The records of referendums and general elections are all gone", and notes how the NLI "will not be able to take the risk of collecting it because of the risk and responsibility that puts on the library in terms of having breached copyright legislation" (Collins, Joint Committee on Tourism, Culture, Arts, Sport and Media, 13 October 2021). Finally, Collins suggests that it

would be useful for the report to go to the Cabinet for consideration and that the report recommend a legislative amendment to copyright legislation, which is the responsibility of the Department of Enterprise, Trade and Employment. That, in time, would allow us to capture those websites and our contemporary history before it is gone forever.

Looking across Europe at our peer national libraries, 60% of European national libraries have this legislation in place and are collecting their countries' websites. We do not want to fall behind and lose the data to a black hole forever (Collins, Joint Committee on Tourism, Culture, Arts, Sport and Media, 13 October 2021).

Dr Collins sums up the situation well and offers more context regarding the urgent need to amend legal deposit legislation for the inclusion of a national domain web archive. From this, we can also surmise that the Department of Tourism, Culture, Arts, Sport and Media is responsible for producing a feasibility report that will outline the necessary requirements to inform the Department of Enterprise, Trade and Employment in the drafting of the necessary copyright/legal deposit legislation.

A more recent example is noted in the NLI Collection Development Policy 2022-2026, where the NLI clearly state their concerns in ensuring that they “collect, as comprehensively as possible, the record of contemporary Ireland”, however this “record is largely online and highly ephemeral.” They conclude that “urgent attention must be given to introducing legislative provision under digital legal deposit for web archiving at scale. Without this, there is a growing and irretrievable gap in the record of Ireland’s history, heritage and recorded knowledge” (NLI, Collection Development Policy 2022-2026).

### 3.4.2 Current status of the NLI web archive

In January 2023, the NLI Web Archive had 3,114 captured websites (since 2011) in the selective public NLI Web Archive (see [Figure 3.7](#)), yet at the end of 2022 there were 329,265 registered domains in the .ie database (.IE Domain Profile Report, 2022, p. 3). Furthermore, that does not account for the thousands of Irish websites registered under other domains such as .com, .net or .org. (see examples in [Table 3.2](#)).

Table 3.2: Examples of Irish websites using .com, .net, and .org

<b>.com</b>	<b>.net</b>	<b>.org</b>
www.irishtimes.com	www.catholicireland.net	www.one-dublin.org
www.tourismireland.com	www.eircom.net	www.ouririshheritage.org
www.corkairport.com	www.fencingireland.net	www.ireland.anglican.org
www.irishexaminer.com	www.fusio.net	www.travelireland.org

It also brings into question the actual parameters being decided for what should be included in an Irish national domain web archive, and who gets to decide? The emphasis seems to be on the inclusion of a .ie domain crawl, but, as has already been shown, basing a national web archive domain solely on a country’s ccTLD, like .ie is too minimal as a representative marker for the collection of a country’s national web estate (Webster, 2019; Day, 2003; Coram, 2015). So, there is a need to evaluate the parameters for what should be included in an Irish national web domain archive from the outset and accounted for in the legislation. For example, websites outside the .ie domain (e.g., .com, .org) could be further demarcated if the website contains an ROI postal address, if the website owner confirms residence at an ROI postal address, if the website lists its place of business as an ROI postal address, if the website is



identified as being hosted on a server located physically in the ROI via a geo-IP lookup, or if the website focuses on the Irish language, or uses a hybrid approach towards English/Gaeilge. Consideration should also be given to requesting permissions from website owners for websites beyond the realms of the ROI web space which reflect the Irish language or if the website belongs to an Irish immigrant or a diaspora community. Websites which have a variety of Creative Commons licences might also be considered for inclusion. Furthermore, the users of web archives should also be involved in the discussion of the selection criteria for what gets included in a national web domain archive, as pointed out by Jatowt et al. (2008).

### 3.4.3 Web archives in the Irish media

In a reflective article titled 'Breaking in to the mainstream', Winters (2017) discusses the role that media and newspapers can play in highlighting the value and importance of web archives, and draws attention to one of the first examples of web archives being mentioned in the UK "news rather than technology pages" (p. 175). Here Winters (2017) refers to an incident which was reported in *The Guardian* newspaper on November 2013, whereby the Conservative party,

deleted more than a decade's worth of speeches from its website. The story was given an added news angle because one of those speeches was by the then Prime Minister David Cameron praising the Internet for 'making more information available to more people' (Winters, 2017, p. 175).

The article in *The Guardian* also noted that the party also took steps to block access to the captured web pages in the Internet Archive's Wayback Machine (Ramesh & Hern, 2013). Although it is worth noting that the Conservative party website had also been archived by the UK Web Archive, and so the missing web content "had been preserved as a part of the national historical record" (Winters, 2017, p. 175). Nonetheless, for Winters (2017), "the media, and newspapers in particular, have an important role to play" in making "the case for the significance of web archives" (p. 175).

In an Irish context, web archives have been mentioned in the Irish mainstream media on a few occasions, but mainly in specialised sections such as technology. Indeed, it is difficult to find examples where web archives entered the discourse of the mainstream "news" until recently, when Hugh O'Connell reported in the *Sunday Independent* how Sinn Féin wiped "years of media statements" from their website, but also how the missing media statements were available in the Wayback Machine. O'Connell fails to mention that the Sinn Féin website had also been archived by the National Library of Ireland multiple times from 2011 to 2022 (Figure 3.10).

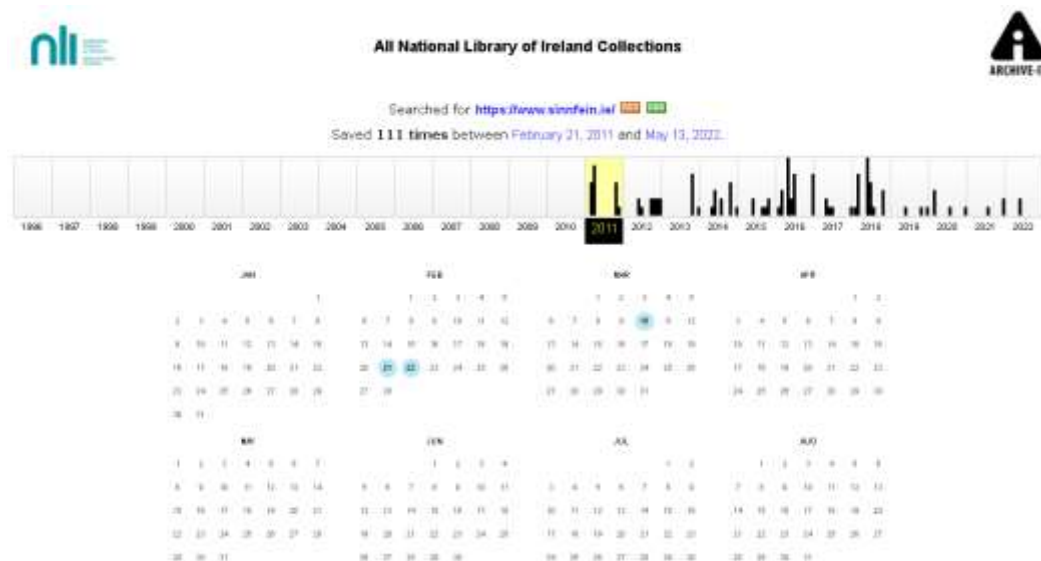


Figure 3.10: Screenshot of NLI Web Archive interface, showing multiple captures of the Sinn Féin website from 2011 to 2022 (www.sinnfein.ie), taken on 2022-10-06

One of the earliest examples of web archives in Irish media is by Michael Cunningham in the ‘COMPUTIMES’ section of *The Irish Times*, and discusses the work of the Internet Archive for the preservation of the world wide web, and he asks: “If a digital national archive is important for the historians of the future, where is Ireland's digital archive?” (Cunningham, 1997a, p. 18). While not referring to web archives specifically, but born digital data in general, in the Business section of *The Irish Times*, Kieran Fagan (2012) discusses the lack of preservational mechanisms in place for Irish born digital heritage, with a focus on the lack of preservation of the digital records of the Irish government in all their manifestations (Fagan, 2012). In 2017, in the Technology section of *The Irish Times*, Charlie Taylor headlined an article with “Ireland's digital content in danger of disappearing, specialist warns”, which discussed the importance of archiving the .ie domain, and how the NLI partnered with the Internet Archive to conduct a crawl of the .ie domain in 2017 (Taylor, 2017b). In 2020, the Librarian and College Archivist of the Library of TCD, Helen Shenton, wrote a Letter to the Editor of *The Irish Times* to discuss the “Digital black hole in our national memory” and the failure of the Irish government to include the web archiving of the Irish national domain as part of the Copyright and Other Intellectual Property Law Provisions Act 2019 (Shenton, 2020).

### 3.4.4 ROI web space in brief

As it stands, the ROI is already “impoverished” (UNESCO, 2003) due to mass losses of digital heritage on the web for the decades of the 1990s, 2000s, and 2010s. It now looks like this will continue well into the 2020s, before the necessary measures are put in place for the collection and preservation of the web space of the twenty-six counties of the ROI in line with the collection and preservation of the web space of the six counties of NI. The parliamentary questions and answers above clearly indicate how the loss of Irish digital heritage to “posterity” is not only due to the lack of a business model for the preservation of digital media as digital heritage and supportive legislation, but also due to political circumstances. Moreover, there is no sense of urgency by the government department responsible for delivering a feasibility report to establish an Irish domain web archive. Even the mainstream media, and the leaders of two Irish political parties seem to be oblivious to the fact that Irish digital heritage is continually haemorrhaging into the digital dustbin. It also adds some meaning to why UNESCO (2003) suggests the loss of digital heritage has often gone unnoticed by societies and nations because “Attitudinal change has fallen behind technological change” and consequently, the economic, social, intellectual, historical, and cultural value, or potential value of the heritage is not realised. It further demonstrates one of the main causes for the loss of Irish digital heritage in the ROI is due to what Lyman (2002) describes as a “cultural problem” whereby “past generations did not, or could not, recognize their historic value” (p. 39). However, surely the Minister responsible for the delivery of a feasibility report on the establishment of an Irish domain web archive realises that the longer they delay in delivering a report, and amending the legislation, they are also ultimately responsible for contributing to the catastrophic loss of Irish digital heritage for current and future generations.

Immediate action is required for an emergency change in the legislation to allow for the collection and preservation of the ROI web estate in the interim, while a feasibility report can continue to be undertaken to advise on the necessary requirements to update the legislation to establish a national web domain archive through “a process of negotiation among interested parties” (Lyman, 2002, p. 40). Moreover, as demonstrated, negotiations should be inclusive of a wide variety of representatives across multiple sectors such as education and teaching, users of web archives for multiple purposes, information professionals with experience in the transition from small-scale selective web archiving to large-scale domain web archiving, and information professionals who are experienced in working with Irish based information ecosystems, and thus, minimise the challenges from the start.

## 4. AWARENESS OF, AND ENGAGEMENT WITH, WEB ARCHIVES IN IRISH ACADEMIC INSTITUTIONS

*If researchers today want to fully understand the present, as well as our past from the mid 1990s onwards, the Web will play a critical role. While there is no common rule for when a topic becomes 'history', the timeframe seems to be shortening as the speed of information dissemination accelerates (Brügger & Milligan, 2019, p. xxviii).*

### 4.1 Introduction

The previous section examined the availability and accessibility of web archives based on the island of Ireland, and their usefulness as resources for Irish based research. In doing so it also examined the causes for the loss of Irish digital heritage and how this contributes to the challenges for web archive research in Ireland. Regarding the establishment of an Irish national web archive domain, the chapter also emphasised the need for the inclusion of multiple representatives in the negotiations for the establishment of an Irish national web archive domain, inclusive of representatives from the teaching and education sectors, users of web archives for multiple purposes, information professionals with experience in the transition from small-scale selective web archiving to large-scale domain web archiving, and information professionals who are experienced in working with Irish based information ecosystems.

As mentioned previously, very little is known about the users or potential users of web archives in Ireland. Indeed, publication of Irish based research integrating the use of archived web content is difficult to find with some exceptions being Malone (n.d.), Harjani (2018), Byrne (2019), Greene & Ryan (2019), Healy (2019), Webster (2019), and Greene (2020). Moreover, to date, and to the best of our knowledge, there have been no web archive user studies conducted across Irish academia which examine scholarly engagement, or awareness of the existence of web archives as resources for research. And, as pointed out by the web archivist at the NLI Web Archive, “‘It’s difficult to get good analytics on web archive users, due to the fact the selective web archive can be accessed remotely’” (Ryan qtd. in Vlassenroot, 2019, p. 100). In essence, very little is known about those who engage with, or might potentially engage with, web archives as resources for Irish based research. Therefore, it is difficult to assess what types of support and incentives will be most effective for assisting

scholars and educators in the use of the archived web for Irish based research. This presented an opportunity for further investigation.

Utilising an online survey, this section investigates awareness of web archives, and engagement or non-engagement with web archives by lecturers, researchers, and students in Irish academic institutions.

The main objectives for the study are outlined below:

- To investigate the awareness of web archives and archived web content as a resource for study/research,
- To generate a better understanding of the users of web archives in Irish academic institutions, and how and why archived web content is used or not used for study/research,
- To explore the challenges and opportunities for using web archives and archived web content as a resource for study/research.

In pursuit of these aims, the survey focuses on the following research questions:

- What is the current level of awareness for the existence of web archives?
- What are the reasons for a lack of engagement with web archives for research?
- What is the likelihood of a non-user using a web archive for research, after becoming aware of its existence?
- Who are the users of web archives?
- How and why are web archives (and web archived content) used for research?
- What is the perceived value of web archives?
- What is the perceived importance of archiving websites based on specific topics?
- What kind of challenges do scholars perceive for the future use of archived web content in their field of research?

The design of the survey considers the objectives of the study and a review of related literature on web archive user studies with attention to research engagement studies conducted by Costea (2018) and Riley and Crookston (2015). The study is exploratory in terms of assessing some of the challenges and opportunities for using web archives, and considerations for how to best facilitate their use, going forward. The methodology for participant recruitment, survey design, and the preparation of the survey dataset is set out, and the survey results and analysis are discussed. It concludes with a review of the findings in relation to the research questions for the study.

## 4.2 Related Literature

### 4.2.1 Web archive user studies

There have been several web archive user studies conducted to date. Some studies focus on engagement with web archives by users in general (Jatowt et al., 2008; Costa & Silva, 2010; Moiraghi, 2018), or focus on scholarly awareness of web archives, and scholarly research engagement and non-engagement (Hockx-Yu, 2014; Riley & Crookston, 2015; Costea, 2018). Other studies focus on web archiving practices, and by correlation, they also examine how web archive collections are being used (Bailey et al., 2014; Bailey et al. 2016; Farrell et al., 2018), and the challenges for research engagement (Truman, 2016; Vlassenroot et al. 2019). A review of some of this literature is available in [section 2.4](#).

### 4.2.2 Use of web archives for Irish based research

Literature integrating the use of archived web content for Irish based topics or research is difficult to find, with some exceptions being Malone (n.d.), Harjani (2018), Byrne (2019), Greene & Ryan (2019), Healy (2019), Webster (2019), and Greene (2020).

Malone's (n.d.) contribution comes in the form of a web page titled 'Early Irish Web Stuff'. David Malone was a student and system administrator in TCD School of Maths in the early 1990s and was soon introduced to Unix and the set up for TCD TCP/IP internet system (Malone, 2016). Realising he had lived through "interesting times", Malone attempted "to try and record or make notes of some of what had happened" (Malone, 2016). This resulted in the development of a web page, titled 'Early Irish Web Stuff', which attempts to track down some of the early Irish websites, and some of the developments happening at the time. Malone consults newspapers, examines conversations in the Usenet archive in Google Groups, consults legacy web pages such as Déjà vu news and the 'What's New' directory on the Mosaic Communications Corporation, and utilises the Wayback Machine to track down the URLs of early Irish websites and web pages (Malone, n.d.).

Using archived social media, Harjani's MSC thesis entitled 'Investigating Information Sharing Behaviour on Twitter: The Case of the Irish Referendum' explores the factors that affect the information people share and consume online using Twitter data archived around the Thirty-sixth Amendment of the Constitution of Ireland or more commonly known as the Repeal of the Eighth Amendment (Harjani, 2018). In doing so, Harjani uses the repost count on Tweets archived as a dependent variable while content-level features were selected as independent variables. The data was then modelled using simple regression methods and supplementary

network analysis. The main findings show that “negative sentiment is a strong driver of reposts” (p. 2). Harjani (2018) further suggests that:

Conversely, posts by news, campaign and politician accounts do not fare well, exhibiting a negative relationship with repost count. A third finding displays the tendency of campaigners to retweet other campaigns of the same vote endorsement. The consequences map out onto the observed polarisation trend in recent years and the rise of fake news. Some of these findings present evidence in support of present literature posing important theoretical and practical questions for policymakers (Harjani, 2018, p. 2).

Harjani's research highlights the important role that social media can play in understanding key events in Irish society and highlights the need for discussions regarding the preservation of social media as national digital heritage, and its inclusion in legal deposit schemes (Harjani, 2018).

Byrne (2019) uses a literature review and three case studies to examine approaches for studying women's sport history, with a particular reference to women's football (soccer), which has been relatively understudied in comparison to male sport histories. The paper reviews digital research methods and offers three potential approaches for studying women's football history, through using web archives as a source for research, digitised newspaper collections, and oral histories. Of relevance for this study is Byrne's (2019) demonstration of using a web archive using the SHINE interface. SHINE is a prototype of a potential research tool that can be used to access and analyse web archive data. It was developed as part of the Big UK Domain Data for the Arts and Humanities project funded by the UK Arts and Humanities Research Council. The data that underpins this service was acquired by JISC from the Internet Archive and includes all .uk websites in the Internet Archive web collection crawled from 1996 to April 2013, when NPLD came into effect (UK Web Archive, n.d., JISC UK Web Domain Dataset; Byrne, 2019). SHINE allows users to search the dataset and refine their using keywords, date ranges and Boolean search terms. Byrne (2019) further offers some examples of Boolean search terms for investigating the history of women's soccer in Ireland.

Greene & Ryan (2019) use a subset of data from the selective NLI Web Archive, that was captured in 2016 and developed “a manually-curated core set of 299 popular Irish domains, corresponding to over 68 million web pages, stored as 27,400 individual ARC files”. First, they extracted the hyperlinks between all HTML pages in the dataset and looked for link pairs (a source and a target URL). Then they convert each link pair into a domain pair and focus on the domain pairs relating to the core set of 299 domains. They find “a dense core” of the

dataset network consists of media, governmental, and sporting websites, and they conclude with suggestions for future work (Greene & Ryan, 2019).

Healy (2019) discusses the decriminalisation of homosexuality in Ireland in line with developments of the world wide web and explores methodologies for finding and recording early internet and web histories of Irish LGBT+ activism. Healy notes how LGBT+ news and activities originating from Irish based email addresses can be found in the Queer Resource Directory as early as 1992, and posts can be traced to Irish based members in the Usenet soc.motss newsgroup from at least 1991. In relation to the web, Healy discusses how the Wayback Machine is useful for examining websites of Irish LGBT+ organisations from the 1990s onwards and demonstrates how such websites underwent “many transformations not merely due to technology, but in terms of discourse and web content, as a result of changes achieved in the social and legal landscape, and an increasing liberalisation to discuss topics which were previously Taboo” (Healy, 2019). Healy also discusses how the Irish 2015 Marriage Equality Referendum was a climax of a twenty-year campaign, to ensure LGBT+ citizens were afforded equal rights and protections. Nonetheless, multiple campaign websites have since disappeared from the live web. For Healy, this amounts to the loss of Irish social, cultural, political, and constitutional history, and reinforces the need for web archiving as a solution for preventing such losses (Healy, 2019).

Using a case study of a web estate of Christian churches in Northern Ireland (NI), historian Peter Webster (2019) examines the nature of the .uk ccTLD as a proxy for the UK web space. Using publicly available documentation such as directories which list individual parish or congregational websites for Roman Catholic churches, Anglican churches, Presbyterian churches, and Baptist churches, Webster compiled a list of relevant website URLs. Then, using archived web data and hyperlink analysis, Webster (2019) examines the link relationships between NI churches, including “the regional, national and cross-border relationships that they imply” (p. 111). For this, Webster (2019) uses a dataset made available in the UK Host Link Graph, which is derived from a larger dataset, being the JISC UK Web Domain Dataset and is made available by the British Library.<sup>43</sup> In doing so, Webster (2019a) draws attention to the difficulties in delimiting the UK web domain solely using the .uk ccTLD as a proxy, due to the vast amount of website content which exists outside of those parameters. For example, UK websites hosted on .com, or .org (Webster, 2019, p. 112). Webster finds that out of 100 domains for Roman Catholic churches in NI, only 12 were registered with a .uk domain, while

---

<sup>43</sup> Host Link Graph JISC UK Web Domain Dataset (1996-2010),  
<https://data.webarchive.org.uk/opendata/ukwa.ds.2/host-linkage/>



3 were registered with the Republic of Ireland's .ie domain. Webster (2019) finds that the links relationships show a very loose mapping to the UK ccTLD (p. 111), and thus, Webster (2019) suggests that that "for web archivists and scholars alike the ccTLD is a weak proxy indeed for the national web" (p. 120).

Greene (2020) offers a useful demonstration of network analysis using WARC data from the 2018 Irish Presidential Election captured in the public NLI Web Archive. Consisting of 1,000 WARC files, containing 57,065 HTML pages, Greene (2020) extracted links from each page, and mapped each link to a pair of domains, with a focus "on pairs of domains for which both the source and target are distinct", thus excluding internal links. Greene (2020) suggests that

By representing large collections of web pages as a link network, researchers can apply existing methodologies from the field of network analysis. For web archives, we can use these methods to explore their content, potentially identifying meaningful historical trends (Greene, 2020, EWA Book of Abstracts)

Greene highlights how the use of network analysis can benefit the collection development of the NLI selective web archive, as well as for studying the "archived Irish web" (Greene, 2020, EWA Book of Abstracts, 2020).

The literature above provides a useful starting point when considering the type of research that has already been undertaken using web archives for research on Irish based topics, and how it can be built upon. It further demonstrates the use of a qualitative approach (Malone, n.d., Healy, 2019), a big data approach (Greene & Ryan, 2019; Greene, 2020) and combining qualitative and big data approaches (Harjani, 2018; Byrne, 2019; Webster, 2019). This provides a good indicator on the types of research which need to be accounted for in any forthcoming legislation on copyright and legal deposit. Moreover, Harjani's (2018) research highlights the important role that social media can play in understanding key events in Irish society, thus, any new legal deposit legislation introduced in Ireland should consider making provisions for the inclusion of social media content.

## 4.3 Methodology

In this section, the methodological approach for the chapter is laid out, including the design for the online survey, the recruitment process, and the approaches for data collection and analysis. Online (questionnaire) surveys are a research method for gathering information about behaviours, attitudes, values, and experiences across a broad range of research disciplines and can be used as a standalone method or as part of a combined approach (Dawson, 2020, p. 288). As with all research methods, there are advantages and

disadvantages which need to be considered for conducting an online survey as a research method for a user study (Wright, 2005; Steber, 2016). The research for this chapter was conducted in compliance with best practice guidelines for the collection and management of research data, as outlined in Maynooth University Research Ethics Policy (2016) and Maynooth University Research Integrity Policy (2016). To note here these policies were updated in 2019 (Ethics) and 2021 (Integrity), after the data had been collected and analysed. However, this did not affect the research plan. The collected/analysed data will be migrated to a private server repository in Maynooth University, for long-term preservation, for a period of ten years, after which, it will be deleted in full (as outlined in MU Research Integrity Policy, 2021).

### 4.3.1 Survey software

The data was collected using the SurveyMonkey online survey tool. At the time the research was being carried out, Maynooth University did not have a specific policy for conducting online surveys, or a policy on which type of software to use for such studies. Thus, the researcher opted to use SurveyMonkey due to having prior experience in using the software and having an account subscription. To note here, Maynooth University only introduced a policy for online surveys in November 2019, which incorporates the use of JISC Online Surveys tool as the only tool permitted by the university for conducting online survey studies of this nature.

### 4.3.2 Survey recruitment

The survey was accessible via a web link inserted in recruitment emails and several posts on the researcher's social media platforms (Facebook, Twitter, and LinkedIn). From November to December 2018, 970 recruitment emails were sent to academics (mostly head of departments, or head of degree programmes) and to department administrators in all fields of research at nine universities in the Republic of Ireland (ROI) and Northern Ireland (NI). The survey was open from November 2018 until January 2019. [Table 4.1](#) provides the list of universities, and a breakdown of the number of emails sent per university.

As well as providing the survey link, emails provided information on the purpose of the study, with an assurance for anonymity and confidentiality. See [Appendix A](#) for an example of the recruitment email. As a degree of self-selection bias was expected due to the interests of those who are aware of, or use web archives, the recruitment email also emphasised the equal importance of participation from respondents who were not aware of or did not engage with web archives.

Table 4.1: Breakdown of recruitment emails sent per university

Third-Level Academic Institutions	Sent to Academics	Sent to Admins	Total per Uni
Dublin City University	=49	=37	=86
Maynooth University	=54	=42	=96
National University of Ireland, Galway	=77	=37	=114
Queen's University Belfast	=79	=26	=105
Trinity College Dublin	=102	=40	=142
University College Cork	=70	=45	=115
University College Dublin	=116	=33	=149
University of Limerick	=49	=36	=85
Ulster University	=55	=23	=78
Totals	(=651)	(=319)	(=970)

Early in the email recruitment process, it was noticed that a few of the complete surveys had inconsistent responses with regards to the awareness of and use of a web archive. For example, when participants were asked to name any other web archives that they were aware of or engaged with, some respondents provided the names of digital archives or digital libraries. At this point it was decided to provide some additional text in the recruitment email, briefly noting the difference between a web archive, and a digital archive/library.<sup>44</sup> Nonetheless, there were similar instances of inconsistencies in later survey responses, and this will be discussed in more detail further on.

### 4.3.3 Survey design and questions

The design of the research and survey questions considered the aims of the chapter and a review of similar web archive user studies ([section 2.4](#)). An effort was made to ensure the survey was answerable in 8-10 minutes, to increase the chances of completion (Chudoba,

---

<sup>44</sup> Additional Note: A 'web archive' is a resource that captures and preserves websites, blogs, and web pages, and provides access to view such content, long after it has disappeared from the live web. A web archive differs from a digital archive/ library in so far as a web archive only contains archived websites, blogs, and web pages.

2018; CoolTool, 2017; Steber, 2016). The survey was tested by four academic colleagues to ensure the questions were clearly understood, after which, some amendments were introduced to the survey language and layout. A final draft of the research project including information about the project, informed consent, how the data would be collected, managed, and used (see [Appendix B](#)) and a copy of the survey questions (see [Appendix C](#)) was submitted to Maynooth University research ethics committee, and the study received ethics approval [SRESC-2018-083] in October 2018.

The survey consisted of thirty questions, but all respondents were not required to answer every question, this was dependent on whether a respondent was a user or non-user of web archives for their research/studies. The questions contained a mix of dichotomous, trichotomous, and multiple-choice questions (some with options for free text), Likert scales, and an optional open-ended question at the end. All participants were asked to answer some demographic questions based on nationality, age, gender, position, and discipline category. All participants were provided with a description of an online public and dark web archive and were asked to answer questions on awareness of their existence, and whether they used web archives for their research or studies. Based on a Yes or No answer, participants were then directed to a set of questions for users or non-users. Non-user questions focused on reasons for non-engagement, and the likelihood of engagement with web archives in the future. Users were asked questions based on their reasons for using web archives and the web archive resources they engaged with. In the final section, all participants were asked to answer questions on their perceived value of web archives, the importance of archiving websites based on different topics, and the significance of web archives as resources for current, medium, and future use in their field of research. The survey ended with an optional open-ended question to allow participants to comment on their perceptions of the challenges for using web archived content in their disciplines in the future.

It is worth discussing here some of the choices that were made in providing a term for the types of web archive collections being held by some web archiving initiatives. The term “online public web archive” is used to describe a resource “whereby access is available to the general public via the web/internet from any location”; and the term “dark web archive” refers to a resource with no public access or with restricted access “onsite in a designated reading room or Library via an onsite portal.” The term dark domain web archive is used to refer to an archived web domain collection which has no public access or has restrictive access. For example, regarding the NLI, we use the term online public NLI web archive (selective web archive collection) and the NLI dark domain web archive (a collection of domain crawls conducted by the NLI). To note, the domain crawls were conducted in 2007

and 2017, and were at the time speculated to become accessible onsite in the NLI reading room (Taylor, 2017a), but were inaccessible at the time the study was conducted.

The reasons for choosing these terms were first guided by the need to come up with terms that would describe the status of web archives with restrictive access to an unfamiliar audience. We were also guided in some way by the use of the term domain dark archive in the Analytical Access to the Domain Dark Archive (AADDA) project. For example, when describing the AADDA project, Webster (2012) describes it as an 18-month project in the UK which sought to “enhance the sustainability of a substantial dark archive of UK domain websites collected between 1996 and 2010 by the Internet Archive, copies of which were recently acquired by the JISC and are stored at the British Library on their behalf” (Webster, 2012; Analytical Access to the Domain Dark Archive, 2012+). And, when discussing the use of this data, Gorsky (2015) describes the dataset as “a large number of UK domain websites, captured 1996–2010, which is colloquially termed the Dark Domain Archive while technical issues surrounding user access are resolved” (p. 596). Thus, it was considered that archived web collections which are not publicly accessible, or which have restrictive access in place may be referred to as a ‘dark web archive’, and archived web domain collections which are not publicly accessible or have restrictions in place may be termed as a ‘dark domain web archive’. From there, it was formulated that the term could be applied to both legal deposit and non-legal deposit archived web domain collections with restrictive access. At the time, it was felt that the use of these terms would be a reasonable way of describing the context of such collections to an unfamiliar audience.

However, since then, it was reconsidered that the use of those terms may also be somewhat out of line with evolving terminologies and trends in the use of the terms, dark archives (custodian access only), dim archives (mix of dark and open), and open archives (light) (Skinner & Schultz, 2010, pp. 128–131; Erickson, 2013). Although there are examples where these boundaries are blurred. For example, Lavoie and Dempsey (2004) assert that the “notion of ‘dark archives’, supporting little or no access to archived materials, has met with scant enthusiasm in the library community”, and suggests that dark archives “will function not just as guarantors of the long-term viability of materials in their custody” but also offer “access gateways.” In addition, Martzahl (2010) describes a dark archive as “a secret place for storing archival material with restricted user access.”

#### 4.3.4 Survey responses

The survey was open from November 2018 until January 2019. 378 participants responded to the survey through email (=367) and social media (=11). However, 93 participants exited the

survey prior to completion. This amounted to a completion rate of 75.40%. As participants were informed that their responses would not be recorded if they did not complete the survey, the 93 incomplete surveys were removed and deleted. A further 46 complete surveys were also removed from the survey dataset, due to response inconsistencies, and will be discussed next.

As mentioned previously, early in the recruitment stage, it was noticed that some surveys had inconsistent responses with regards the awareness of and use of a web archive, in so far as some respondents confused a web archive with other types of resources such as digital libraries, digital archives, and data repositories such as Project Gutenberg, JStor and Talkbank. In total, there were 28 such instances. As this study was aimed to address the reasons for user and non-user engagement, it was decided not to include the 28 survey responses in the final tally for analysis. This is also comparable to an occurrence in the study conducted by Costea (2018). As Costea's study was specifically aimed at users of archived websites, Costea did not factor inconsistent responses in the results. Moreover, there were further instances of inconsistencies in this survey. 18 respondents identified as a user of web archives, yet they indicated that they were not aware of and did not use any of the web archives that were listed, nor did they provide a name of any other web archive they were aware of or used. Riley and Crookston (2015) also came across a similar occurrence in their study of academic institutions; however, they opted to include this data for final analysis, but used filters to calculate their results around the inconsistencies. For this study, however, it was decided not to factor in the 18 surveys with such inconsistencies, to provide a clearer representation for users and non-users and the reasons for engagement or non-engagement with web archives. Therefore, the final tally of complete surveys for analysis in this chapter is (N=239).

#### 4.3.5 Survey limitations

Participation was voluntary, and participants could withdraw at any time during the process of filling out the questionnaire, with the knowledge that their responses would not be recorded. It is also worth noting that some fields of research are more dominantly represented in some universities than others, and some fields of research are not equally available across all universities. Consequently, this may have resulted in an over-representation of participants from some fields of research. It is not possible to evaluate this effect due to ethics considerations, as there were no identifiers collected to evaluate a response rate per university/department. Nonetheless, in the final tally of survey responses for analysis, the respondents identified with twenty-four discipline categories, providing a varied range of representations from different fields of research. Also, as with all studies

based on survey sampling, this survey study cannot be construed to represent the academic population in Ireland as a whole.

## 4.4 Results and Analysis

The survey results and analysis are based on a final number of survey respondents (N=239), and percentages in the discussion and graphs are reflective of this, unless otherwise stated in the case of user and non-user questions and answers. In this instance, 59 respondents identified as a user of web archives, and 180 respondents identified as a non-user.

### 4.4.1 Demographics

This section provides an overview of responses to questions on nationality, age, gender, position, and discipline categories with some data breakdowns for representations of users and non-users. The purpose of these questions was to establish some demographic information about participants which might reveal some trends when cross-tabulated with data from other responses in the survey.

#### 4.4.1.1 Nationality, age, gender

Respondents (N=239) identified with 20 nationalities. [Table 4.2](#) offers a representation of participant responses for nationality (N=239), with a comparison of nationality representations for users and non-users. As expected, the highest rate for identification with a nationality was Ireland (75.73%, n=181); of which 44 respondents identified as a user and 137 as a non-user. [Table 4.3](#) provides an overview of participant responses for age, with a comparison of age representations for users and non-users. Of overall participation (N=239), the highest representation for age is the age bracket of 45-54 (24.27%, n=58) with 14 respondents identifying as a user. This is followed by the age brackets of 18-24 (21.76%, n=52), and 35-44 (20.50%, n=49). Of overall participation (N=239), there were slightly more female (52.30%, n=125) respondents than male respondents (45.61%, n=109). [Table 4.4](#) provides a comparison of user and non-user gender representations.

Table 4.2: Representation of participant responses for nationality (N=239), with a comparison of nationality representations for users and non-users

Nationality Answer Choices	Responses (N=239)		User (n=59)	Non-User (n=180)
AT - Austria	0.42%	(n=1)		=1
AU - Australia	0.42%	(n=1)		=1
BG - Bulgaria	0.84%	(n=2)		=2
CA - Canada	0.42%	(n=1)		=1
DE - Germany	3.35%	(n=8)	=5	=3
ES - Spain	1.67%	(n =4)	=1	=3
FR - France	0.84%	(=2)		=2
GB - United Kingdom	5.02%	(=12)	=3	=9
IE - Ireland	75.73%	(=181)	=44	=137
IN - India	1.67%	(n =4)		=4
IT - Italy	2.51%	(n =6)		=6
MW - Malawi	0.42%	(n =1)	=1	
NG - Nigeria	0.42%	(n =1)		=1
NL - Netherlands	0.42%	(n =1)		=1
RO - Romania	0.42%	(n =1)		=1
RS - Serbia	0.42%	(n =1)		=1
RU - Russia	0.42%	(n =1)	=1	
SI - Slovenia	0.84%	(n =2)		=2
UA - Ukraine	0.42%	(n =1)	=1	
US - United States	3.35%	(n =8)	=3	=5



Table 4.3: Representation of participant responses for age (N=239), with a comparison of age representations for users and non-users

Age Bracket Answer Choices	Responses (N=239)		User (n=59)	Non-User (n=180)
18-24	21.76%	(n=52)	=11	=41
25-34	19.67%	(n=47)	=10	=37
35-44	20.50%	(n=49)	=12	=37
45-54	24.27%	(n=58)	=14	=44
55-64	11.72%	(n=28)	=11	=17
65+	1.26%	(n=3)	=1	=2
Prefer not to say	0.84%	(n=2)	=0	=2

Table 4.4: Representation of participant responses for gender (N=239), with a comparison of gender representations for users and non-users

Gender Answer Choices	Responses (N=239)		User (n=59)	Non-User (n=180)
Male	45.61%	(n=109)	=34	=75
Female	52.30%	(n=125)	=24	=101
Other	1.26%	(n=3)	=0	=3
Prefer not to say	0.84%	(n=2)	=1	=1

#### 4.4.1.2 Positions and disciplines

To better understand the positions of participants within academic institutions, they were provided with a choice of seven position categories or the option of 'Other' to enter free text. 16 respondents provided free text which was coded, of which 10 responses were incorporated into the existing categories that were offered.<sup>45</sup> The six remaining responses were coded into three new categories which are marked with an asterisk (\*). These

---

<sup>45</sup> 6 respondents identified as a lecturer to some degree, such as 'Part time lecturer' or 'Lecturer above the bar' - these were added to the existing category for Senior Lecturer or Associate Lecturer; 1 respondent identified as a 'PhD candidate and teaching fellow' - this was added to the existing category of PhD candidate/student; 3 respondents identified as 'Assistant Professor' - these were added to the existing category for Professor or Associate Professor.

adjustments are included in the final calculations. [Table 4.5](#) provides an overview of participant responses for position, in line with representations for users and non-users. [Figure 4.1](#) provides a representational graph of the position of users in line with total responses (N=239) and shows that users of web archives within this study are educators (12.55%, =30), researchers (7.95%, =19), and students (4.18%, =10).<sup>46</sup>

In configuring a research/discipline area, participants were offered a choice of 20 categories representing a discipline or collective of disciplines, and the option of 'Other' to enter free text. 29 respondents chose 'Other' and entered free text of which was coded and added to existing categories or incorporated into seven new categories which are marked with an asterisk (\*). These adjustments are included in the final calculations.

Table 4.5: Representation of participant responses for position (N=239), with a comparison of position representations for users and non-users

Position Answer Choices	Responses (N=239)		User (n=59)	Non-User (n=180)
Undergraduate student	20.08%	(n=48)	=8	=40
Postgraduate student	10.46%	(n =25)	=4	=21
PhD candidate/student	15.90%	(n =38)	=7	=31
Postdoctoral associate, researcher or fellow	7.53%	(n =18)	=4	=14
Employed researcher in a third-level educational setting or project	4.18%	(n =10)	=6	=4
Senior Lecturer or Associate Lecturer	21.34%	(n=51)	=16	=35
Professor or Associate/Assistant Professor	17.99%	(n =43)	=14	=29
*Administrator (academics/research)	1.26%	(n =3)	=0	=3
*Technical/Support Staff	0.84%	(n =2)	=0	=2
*Director of research centre	0.42%	(n =1)	=0	=1

---

<sup>46</sup> Figure 4.1 - Representational Graph: Educator (Senior Lecturer/Associate Lecturer + Professor/Associate Professor); Student (Undergraduate + Postgraduate + PhD candidate/student); Researcher (Postdoctoral associate, researcher, or fellow + Employed researcher in a third-level educational setting or project).

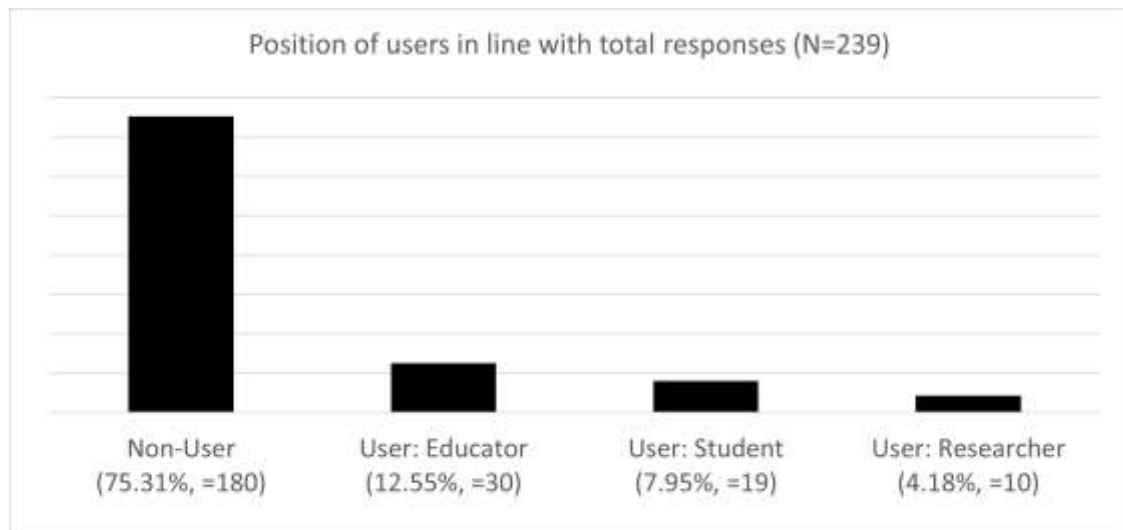


Figure 4.1: Position of users (n=59) under the representations of educators, researchers, and students, in line with total responses (N=239)

Respondents identified with 24 discipline categories. [Table 4.6](#) offers a breakdown of responses for a discipline category in line with user and non-user representations. User respondents represent 17 different discipline categories. The number of users vis-à-vis the rate of participation per discipline category shows a strong number of users from the Humanities (21 of 50) but a low number of users from Social Sciences (5 of 33), Engineering Science (3 of 24), and Natural Sciences (2 of 29).

Table 4.6: Representation of participant responses for discipline category (N=239), with a comparison of discipline representations for users and non-users

Discipline Answer Choices	Responses (N=239)		User (n=59)	Non-User (n=180)
Architecture	0.84%	(n=2)	=1	=1
Arts (visual, performance, music)	1.26%	(n=3)	=0	=3
Business/Economics/Finance	2.51%	(n=6)	=3	=3
*Built Environment	0.42%	(n=1)	=0	=1
Computer Science	4.60%	(n=11)	=3	=8
*Construction Management	0.42%	(n=1)	=0	=1
*Dental Science	0.42%	(n=1)	=1	=0
Digital Arts/Humanities/Cultural Heritage	1.67%	(n=4)	=1	=3
Educational Science	5.44%	(n=13)	=3	=10
Engineering Science	10.04%	(n=24)	=3	=21

Geography (cartography, hydrology, meteorology, environment)	1.67%	(n=4)	=2	=2
Government/Public Administration	0.42%	(n=1)	=1	=0
*Health Studies/Sciences	4.60%	(n=11)	=0	=11
Heritage Studies, Archival Studies	0.42%	(n=1)	=0	=1
Humanities (history, archaeology, languages, literature, philosophy, theology)	20.92%	(n=50)	=21	=29
Internet Studies	0.00%	(n=0)	=0	=0
Law (criminal, civil, common, statute)	7.53%	(n=18)	=6	=12
Library and Information Sciences	0.00%	(n=0)	=0	=0
Mathematics	1.67%	(n=4)	=1	=3
*Medicine/Biomedical Engineering	0.84%	(n=2)	=0	=2
Media/Communications	1.67%	(n=4)	=3	=1
Natural Sciences (biology, chemistry, physics, earth sciences, space sciences)	12.13%	(n=29)	=2	=27
*Nursing/Midwifery	3.77%	(n=9)	=1	=8
Political Science	2.51%	(n=6)	=2	=4
*Psychotherapy	0.42%	(n=1)	=0	=1
Social Sciences (anthropology, human geography, linguistics, sociology, psychology)	13.81%	(n=33)	=5	=28
Sport and Leisure	0.00%	(n=0)	=0	=0

#### 4.4.2 Engagement with online digital-based resources

Provided with four options of online digital-based resources (World Wide Web, Digital Archives, Digital Libraries, and Virtual Research Environments), participants were asked to indicate the frequency to which they use or access these resources for their research or studies. Responses differed for each resource type, but there was a clear indication that 88.28% (n=211) of respondents 'Always', and 11.30% (n=27) 'Sometimes' use the World Wide Web, suggesting that the web is a major research resource in Irish third-level academic institutions. [Table 4.7](#) provides a breakdown for each resource.

Table 4.7: Representation of participant responses for engagement with other online/digital resources

FREQUENCY	World Wide Web (N=239)	Digital Archives (N=239)	Digital Libraries (N=239)	Virtual Research Environments (N=239)
Always	88.28% (=211)	25.10% (=60)	39.33% (=94)	3.77% (=9)
Sometimes	11.30% (=27)	27.20% (=65)	35.15% (=84)	14.64% (=35)
Rarely	0.42% (=1)	20.92% (=50)	12.97% (=31)	28.45% (=68)
Never	0.00% (=0)	26.78% (=64)	12.55% (=30)	53.14% (=127)

### 4.4.3 Awareness of the existence of web archives

This section provides an overview of responses for awareness of the NLI web archive (online public and dark web archive), awareness of other international public web archives, and awareness of any other web archives not mentioned. There is also some cross-tabulation of the data with position and discipline representations.

#### 4.4.3.1 Awareness of the NLI Web Archive

Provided with a ‘Yes’ or ‘No’ option, participants were asked if they were aware that the NLI archives websites which are made accessible through the online public NLI Web Archive (<https://archive-it.org/home/nli>). In addition, they were also asked if they were aware that the NLI archived the Irish domain (.ie) in 2007 and 2017 and would soon make it available as a dark (.ie) web archive – only accessible onsite in a designated reading room at the NLI (Taylor, 2017a; also see NLI, n.d., Irish Domain Web Archive). To note here, at the time this survey was conducted, there was a belief that these collections would soon become accessible in the NLI reading room, as is noted by Charlie Taylor in *The Irish Times* (Taylor, 2017a). However, as was observed in [section 3.4](#) these collections remain inaccessible due to legalities.

Of all respondents (N=239), 18.41% (n=44) indicated that they were aware of the online public NLI Web Archive and identified their nationalities as Ireland (=33), United States (=4), Germany (=2), and United Kingdom (=1). The 44 respondents identified with 11 different discipline categories, of which Humanities (=21) was the most represented. However, there was a low-level of awareness of the resource apropos the rate of participation from respondents from the Social Sciences, Law, Natural Sciences, and Engineering Science (see [Table 4.8](#) for a breakdown). For example, of total participation, 33 respondents identified with the Social Sciences, but only 4 were aware of the online public NLI Web Archive.

Table 4.8: Representation of a comparison of discipline categories of respondents who indicated awareness of the online public NLI Web Archive

Discipline category of respondents	Number of respondents who indicated awareness of the public NLI Web Archive (n=44)	Total number of user/ non-user respondents who identified with that discipline category (N=239)
Business, Economics, Finance	=1	=6
Computer Science	=2	=11
Digital Arts/Humanities/Heritage	=3	=4
Educational Science	=2	=13
Engineering Science	=2	=24
Geography	=2	=4
Humanities	=21	=50
Law	=4	=18
Media/Communications	=2	=4
Natural Sciences	=1	=29
Social Sciences	=4	=33

In the case of the NLI dark (.ie) web archive, only 2% (4 of 239) of respondents indicated that they were aware of its existence. The 4 respondents identified their nationality as Ireland; their discipline categories as Humanities (=3), and Law (=1); and their positions as Lecturer (=2), PhD candidate (=1), and Postdoctoral associate/ fellow (=1).

#### 4.4.3.2 Awareness of other online public web archives

Regarding awareness of other online public web archives, participants (N=239) were provided with a list of six international online public web archives (with the URL link to each resource), and options for a 'Yes' or 'No' answer. [Table 4.9](#) provides a breakdown of responses for each resource and shows a highest degree of awareness for the Internet Archive, Wayback Machine (31.38%, n=75); followed by the US Library of Congress Web Archive (23.85%, n=57); and the UK Web Archive (21.76%, n=52).

Participants were also asked if there were any other web archives (not listed above) that they were aware of and offered an option to enter free text. 11 respondents provided free text, and their comments are summarised below.

- While acknowledging that it was not the same as a web archive, 1 respondent mentioned the revision histories in Wikipedia: "Not exactly an archive, but Wikipedia

does preserve accessible records of page revisions with data concerning who edited pages and why. This has been important for my research as people sometimes use this as a way to put information into the public domain that the public might not otherwise know to query” (User, Dental Science).

- 2 respondents suggested Google Cache, as a means to retrieve an older version of a website.
- 2 respondents mentioned electoral/referendum collections in the NLI Web Archive.
- 1 respondent suggested the Austrian National Library Web Archive.
- 1 respondent referred to the web archive of the Bibliothèque nationale de France
- 1 respondent noted Archive Team and Zone-H.
- 1 respondent mentioned Archive.is.
- 1 respondent referred to the List of Web archiving initiative Wikipedia page.
- 1 respondent added: ‘If this helps...it is not a web archive, but a digital archive I am using: [https://curia.europa.eu/en/content/juris/c2\\_juris.htm](https://curia.europa.eu/en/content/juris/c2_juris.htm)’ (Non-user, Law).

Table 4.9: Representation of participant responses (N=239) for awareness of other online public web archives

Other online public web archives	Responses (N=239)	
	Yes: I was aware	No: I was not aware
Internet Archive, Wayback Machine	31.38% (n=75)	68.62% (n=164)
PRONI Web Archive	13.81% (n=33)	86.19% (n=206)
UK Web Archive	21.76% (n=52)	78.24% (n=187)
UK Government Web Archive	15.06% (n=36)	84.94% (n=203)
UK Parliament Web Archive	12.97% (n=31)	87.03% (n=208)
US Library of Congress Web Archive	23.85% (n=57)	76.15% (n=182)

#### 4.4.4 Engagement with web archives for personal and research interests

To initiate inquiry into the reasons for the use of, or non-use of web archives for research, all participants (N=239) were asked: (i) if they ever accessed or used an online public web archive for their personal interest; and (ii) if they ever accessed or used an online public web archive, or dark web archive to assist with their studies or research. Regarding the use of an online web archive for personal interests (N=239): 30.96% (n=74) indicated 'Yes'; 38.08% (n=91) indicated 'No'; and 30.96% (=74) indicated 'Unsure' (Figure 3.3). For the use of web archives for their studies or research (N=239): 24.69% (=59) indicated 'Yes', and 75.31% (n=180) indicated 'No' (Figure 4.2). Respondents were then directed to corresponding sections for users and non-users.

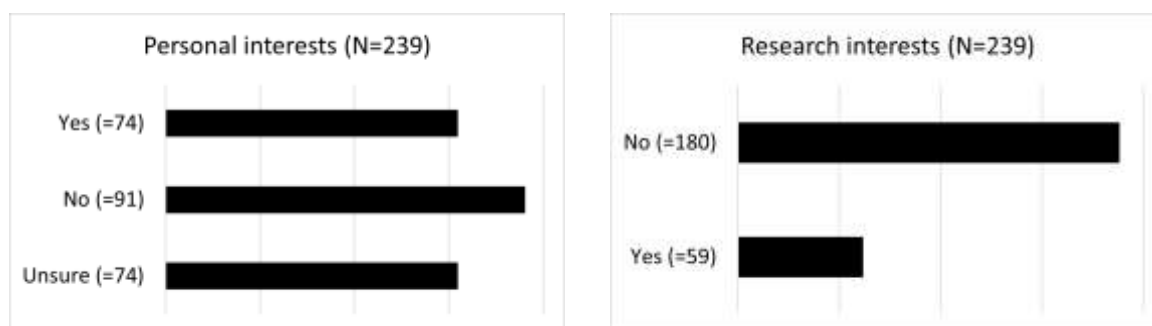


Figure 4.2: Representation of participant engagement with web archives for personal interests and research

#### 4.4.5 Non-users of web archives for research

This section provides an overview of responses by respondents who identified as a non-user (n=180) on questions related to the reasons for a lack of engagement with web archives, and the likelihood of future engagement with online public web archives and a dark web archive.

##### 4.4.5.1 Reasons for lack of engagement with web archives

Respondents (n=180) who indicated that they did not access/use web archives for their research/studies were first asked about their reasons for not using an online public web archive for their studies or research. Participants were provided with seven answer choices, an option of 'Other' to enter free text and were asked to tick all that applied. Table 4.10 provides a breakdown of non-user responses (n=180) and shows that a large majority (78.33%, =141) do not engage with web archives for their research, due to a lack of awareness of the existence of web archives. Other reasons include a lack of knowledge in how to use a web archive (41.67%, =75 ); how to find archived websites in a web archive that are relevant to a research area (45.00%, =81); uncertainty of the credibility or authority of using archived



websites as a primary source (26.11%, =47); and how to cite/reference an archived website from a web archive (17.78%, =32).

26 respondents ticked 'Other reason(s)' for not using an online web archive, and 24 respondents provided free text responses and are summarised as follows,

- 8 respondents indicated that they were unsure as to how relevant, useful, or beneficial, a web archive would be for their research.
- 16 respondents indicated that a web archive was not relevant for their research, of which 4 noted their research required up-to-date sources, and 3 identified as early to modern period historians that required alternate archival sources.

Table 4.10: Representation of non-user respondent (n=180) reasons for not using an online web archive for their studies/research

Answer Choices	Non-User Responses (n=180)
I was not aware of the availability of web archives as resources for my studies/research	78.33% (=141)
I do not know how to use a web archive for my studies/research	41.67% (=75)
I feel that I do not have the technical skills to use a web archive for my studies/research	7.78% (=14)
I do not know how to find archived websites relevant to my studies/research in a web archive	45.00% (=81)
I do not know how to cite/reference an archived website from a web archive to include in my studies/research	17.78% (=32)
I am unsure of the credibility or authority of using archived websites as a primary source for my studies/research	26.11% (=47)
I am unsure about copyright implications for using archived web content for my studies/research	13.33% (=24)
Other reason(s) for not using an online web archive for your studies/ research (please specify)	14.44% (=26)

There is no denying that web archives are simply not relevant for some fields of research. On the other hand, the findings suggest that the value of web archives as a research resource is not clearly understood by an unfamiliar audience. For example, for some respondents, there is a need for more efforts to demonstrate the importance of archiving the web and to promote the value of web archives for research. It could be further suggested that there is a need for the dissemination of use cases in Irish based research that will demonstrate theoretical and methodological approaches for using web archives as a research resource. Of additional interest are the discipline categories of the non-user respondents (=141) who identified with a lack of engagement with web archives, due to a lack of awareness. As mentioned previously, it was surprising to find a low-level of NLI web archive users who identified with the Social Sciences (4 of 33). However, a high-level of respondents from the Social Sciences (22 of 33) and Natural Sciences (20 of 29) do not engage with web archives for research due to a lack of awareness of their existence. A similar case could be made for respondents who identified with some other discipline categories. Table 4.11 offers a breakdown of this data vis-à-vis discipline category.

Table 4.11: Representation of discipline categories for non-user respondents who indicated a lack of research engagement with online web archives due to a lack of awareness (n=141), in line with the total number of user and non-user participants who identified with that discipline category

Discipline categories of non-user respondents, who indicated a lack of research engagement with public web archives, due to a lack of awareness	Number of respondents who indicated non-engagement, due to a lack of awareness (n=141)	Total number of user and non-user respondents who identified with that discipline category (N=239)
Architecture	=1	n=2
Arts	=3	n=3
Business, Economics, Finance	=3	n=6
Computer Science	=7	n=11
Digital Arts/Humanities/Heritage	=2	n=4
Educational Science	=10	n=13
Engineering Science	=17	n=24
Geography	=2	n=4
Health Studies/Sciences	=9	n=11
Heritage Studies, Archival Studies	=1	n=1
Humanities	=18	n=50
Law	=10	n=18

Mathematics	=2	n=4
Medicine, Biomedical Engineering	=2	n=2
Media/Communications	=1	n=4
Natural Sciences	=20	n=29
Nursing, Midwifery	=7	n=9
Political Science	=3	n=6
Psychotherapy	=1	n=1
Social Sciences	=22	n=33

#### 4.4.5.2 Likelihood of future engagement with web archives

Using a Likert scale for answer options, non-user participants (n=180) were asked about their likelihood of using the public NLI Web Archive in the future for their research, as well as some other online public web archives. Participants were provided with a list of six other online web archives (with a URL link to each resource). [Table 4.12](#) provides a breakdown of responses.

Table 4.12: Representation of non-user responses (n=180) for the likelihood of future engagement with online public web archives

Participation per resource (n=180)	Definitely Likely	Fairly Likely	Unsure	Not Very Likely	Definitely Not Likely
NLI Web Archive	11.67% (=21)	30.00% (=54)	17.22% (=31)	28.33% (=51)	12.78% (=23)
Internet Archive, Wayback Machine	8.89% (=16)	26.67% (=48)	23.89% (=43)	23.89% (=43)	16.67% (=30)
PRONI Web Archive	3.89% (=7)	11.67% (=21)	20.56% (=37)	33.89% (=61)	30.00% (=54)
UK Web Archive	4.44% (=8)	24.44% (=44)	20.00% (=36)	27.22% (=49)	23.89% (=43)
UK Government Web Archive	2.78% (=5)	13.33% (=24)	20.00% (=36)	36.11% (=65)	27.78% (=50)
UK Parliament Web Archive	1.67% (=3)	11.11% (=20)	18.33% (=33)	33.89% (=61)	35.00% (=63)
US Library of Congress Web Archive	3.89% (=7)	16.11% (=29)	18.89% (=34)	32.22% (=58)	28.89% (=52)

From there, it is possible to calculate some measurements using filters, for a probability on whether awareness increases the likelihood of research engagement for each resource, by using the following formula.

Formula: number of participants who were unaware of a public web archive resource at the start of the survey, who also identified as a non-user\*\*\* and who specified as definitely likely\* and fairly likely\*\* to use the resource for future research ( $* + ** = [] \div *** \times 100 = [] \%$ ).

Table 4.13 provides an overview of the application of the formula for each resource. It demonstrates a probability percentage, that awareness increases the likelihood for future research engagement with online public web archives for non-user respondents who were unaware of the existence of web archives prior to participation in the survey.

Table 4.13: Representation for the probability that awareness increases likelihood of engagement with online public web archives for non-users (n=180) who were unaware of the existence of online public web archives

Web archive resources	Unaware & non-user***	Definitely Likely*	Fairly Likely**	Increased Likelihood
NLI Web Archive	=162	=16	=46	38.27%
Internet Archive, Wayback Machine	=138	=7	=30	26.81%
PRONI Web Archive	=162	=5	=12	10.49%
UK Web Archive	=161	=6	=34	24.84%
UK Government Web Archive	=169	=4	=19	13.61%
UK Parliament Web Archive	=174	=3	=16	10.92%
US Library of Congress Web Archive	=163	=5	=18	14.11%

Participants were also asked about the likelihood that they would access or use a dark web archive in the future for their studies or research. They were informed that a dark web archive is only accessible onsite in a designated reading room or Library via an onsite portal. Figure 4.3 provides a breakdown of responses, calculated from participation in this section (n=180). As one can see, 5.00% (=9) of non-users responded with 'Definitively Likely', 12.22% (=22) with 'Fairly Likely', and 18.89% (=24) with 'Unsure'. While this seems like a low response towards the likelihood of using a dark web archive in the future, one needs to account that many of these respondents (=141 of 180) indicated their reasons for not using a web archive for research or study was due to a lack of awareness of their existence (see Table 4.10).

Moreover, from the findings in an earlier question ([section 4.4.3](#)), of the total number of respondents in the survey, only 2% (4 of 239) indicated that they were aware of the existence of the NLI dark (.ie) web archive, meaning the concept of a dark (domain) web archive and its value as a research resource may not be clearly understood.

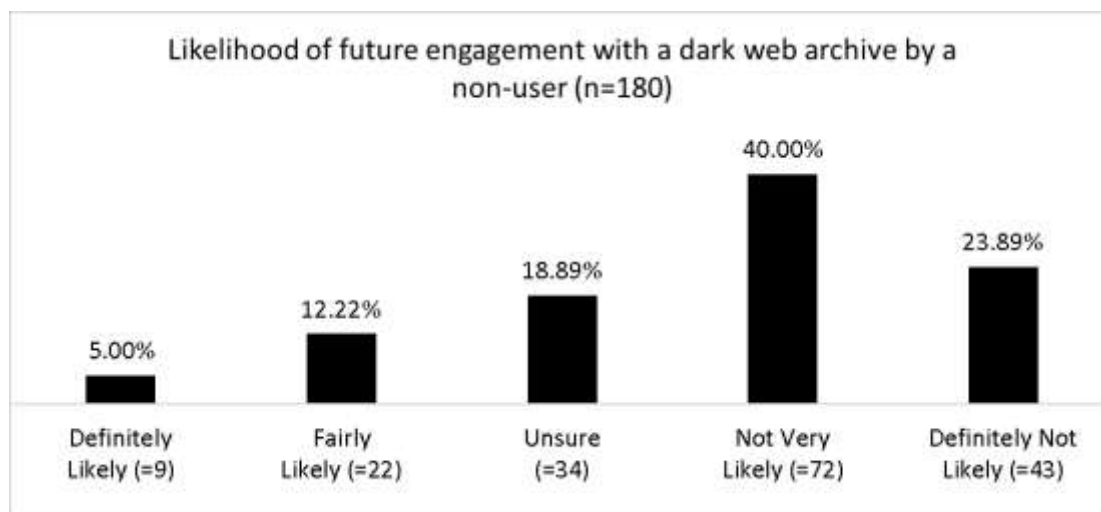


Figure 4.3: Representation for the likelihood of future engagement by a non-user (n=180) with a dark web archive

## 4.4.6 User Engagement with web archives

This section provides an overview of responses by user respondents (n=59) on their general reasons for using a web archive, their reasons for using a web archive for research, their use of online public web archives and a dark web archive, and the likelihood they would use a dark web archive in the future.

### 4.4.6.1 Disciplines of user respondents

The discipline categories of user respondents (n=59) are outlined in [Table 4.14](#) and indicate that user respondents identify with a broad range of research fields.

Table 4.14: Representation of discipline categories for user respondents (n=59)

Discipline Answer Choices		User (n=59)
1	Architecture	=1
2	Business/Economics/Finance	=3
3	Computer Science	=3
4	Dental Science	=1
5	Digital Arts/Humanities/Cultural Heritage	=1
6	Educational Science	=3
7	Engineering Science	=3
8	Geography (cartography, hydrology, meteorology, environment)	=2
9	Government/Public Administration	=1
10	Humanities (history, archaeology, languages, literature, philosophy, theology)	=21
11	Law (criminal, civil, common, statute)	=6
12	Mathematics	=1
13	Media/Communications	=3
14	Natural Sciences (biology, chemistry, physics, earth sciences, space sciences)	=2
15	Nursing/Midwifery	=1
16	Political Science	=2
17	Social Sciences (anthropology, human geography, linguistics, sociology, psychology)	=5

#### 4.4.6.2 General reasons for using a web archive

Respondents who identified as a user (n=59) were asked about their access or use of a web archive in general. Participants were provided with seven answer choices based on interests, along with an option of 'Other' to enter free text. They were asked to tick all that applied. [Figure 4.4](#) provides an overview of participant responses. It indicates that the vast majority (93.22%, =55) of user respondents utilise a web archive for research interests, followed by personal interests (72.88%, =43), historical interests (64.41%, =38), and cultural interests (42.37%, =25). One respondent chose the option for 'Other' and noted using a web archive to "recover old advertisements for teaching" (User, Lecturer, Business, Economics, Finance).

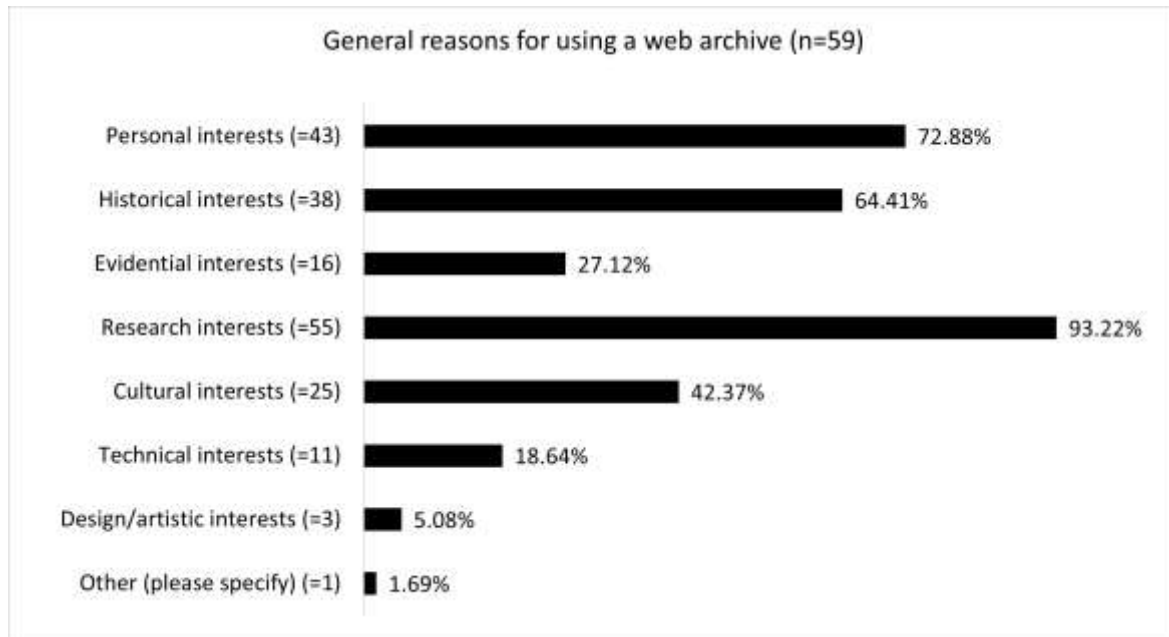


Figure 4.4: Representation of general reasons for using a web archive (n=59)

#### 4.4.6.3 Reasons for using archived web content

Regarding the use of archived web content, user respondents (n=59) were asked about their use of archived web content (archived websites, blogs, web pages). Participants were provided with 6 answer choices, and an option for 'other reasons' to enter free text. They were asked to tick all that applied. [Table 4.15](#) offers a breakdown of responses of participant reasons for using web archived content for their studies/research.

Further to this, 9 respondents entered free text which is summarised below.

- 3 respondents noted using a web archive for personal and historic interests,
- 1 respondent mentioned using archived web content as a secondary source for a thesis,
- 3 respondents referred to accessing content/websites no longer available on the web, with 1 respondent specifying "technical articles" (User, Engineering Science),
- 1 respondent mentioned access to "policy pages that were no longer publicly accessible, so as to compile evidence" (User, Humanities)
- 1 respondent used archived web content for a study of "longitudinal data concerning water parameters [as] sometimes websites only display current or recent results" (User, Dental Science) .

To break this down further, the reasons for using web archives for study or research by user respondents can be further organised by the following themes: for coursework purposes, for professional publication and historical research purposes, for teaching purposes, for

qualitative and quantitative research purpose, and for access to materials no longer available on the live web (Table 4.16).

Table 4.15: Representation of user participant reasons for using web archived content for their studies/research

Answer Choices	User Responses (n=59)
I have used archived web content as a primary source in an academic essay/assignment for my course	42.37% (=25)
I have used archived web content to document the history of an organisation in an academic essay/assignment for my course	20.34% (=12)
I have used archived web content as a primary source in a professional research report	16.95% (=10)
I have used archived web content as a primary source in a professional publication	27.12% (=16)
I have used archived web content to document the history of an organisation in a professional report/publication	11.86% (=7)
I have used archived web content as part of my teaching materials for undergraduate students	18.64% (=11)
I have used archived web content as part of my teaching materials for postgraduate students	22.03% (=13)
I have used large volumes of archived web content for content analysis / textual analysis / discourse analysis	8.47% (=5)
I have used large volumes of archived web content for data mining / topic modelling / data visualisation	6.78% (=4)
I have used large volumes of archived web content for network analysis / geo-spatial analysis	5.08% (=3)
I have used archived web content for other reasons not listed above - please specify	15.25% (=9)



Table 4.16: Representation of user respondent reasons for using web archives for study or research (n=59)

For coursework purposes (=38)	
<ul style="list-style-type: none"> <li>• as a primary source in an academic essay/assignment for my course (=25)</li> <li>• to document the history of an organisation in an academic essay/assignment for my course (=12)</li> <li>• as a secondary source for a thesis (=1)</li> </ul>	=38
For professional publication and historical research purposes (=33)	
<ul style="list-style-type: none"> <li>• as a primary source in a professional research report (=10)</li> <li>• as a primary source in a professional publication (=16)</li> <li>• to document the history of an organisation in a professional report/publication (=7)</li> </ul>	=33
For teaching purposes (=25)	
<ul style="list-style-type: none"> <li>• as part of my teaching materials for undergraduate students (=11)</li> <li>• as part of my teaching materials for postgraduate students (=13)</li> <li>• to recover old advertisements for teaching (=1)</li> </ul>	=25
For qualitative and quantitative research purposes (=12)	
<ul style="list-style-type: none"> <li>• for content analysis / textual analysis / discourse analysis (=5)</li> <li>• for data mining / topic modelling / data visualisation (=4)</li> <li>• for network analysis / geo-spatial analysis (=3)</li> </ul>	=12
For access to materials no longer available on the live web (=5)	
<ul style="list-style-type: none"> <li>• for accessing content/websites no longer available on the web (=3)</li> <li>• for access to “policy pages that were no longer publicly accessible, so as to compile evidence” (=1)</li> <li>• for a study of “longitudinal data concerning water parameters” (=1)</li> </ul>	=5

#### 4.4.6.4 Use of online public web archives for studies/research

In terms of using online public web archive resources, user respondents (n=59) were asked if they ever accessed or used the online public NLI Web Archive and six other online public web archives for their studies or research. [Table 4.17](#) provides a breakdown of responses by user respondents (n=59) for both questions and shows more than half of user respondents indicated being a user of the Internet Archive, Wayback Machine (=35). This was followed by the NLI Web Archive (=23), the UK Web Archive (=22), the UK Government Web Archive (=19) and the US Library of Congress Web Archive (=14). In examining the position and disciplines of respondents who use the public NLI Web Archive (=23 of 59), it reveals it is used by

educators (=9), students (=10) and researchers (=5) from nine different fields of research. A full breakdown of position representations, in line with discipline categories is available in [Appendix D](#) (Table D.1).

Table 4.17: Representation for the use of online public web archive by user respondents

Use of online public web archive resources	User Responses (n=59)
NLI Web Archive	=23
Internet Archive, Wayback Machine	=35
PRONI Web Archive	=5
UK Web Archive	=22
UK Government Web Archive	=19
UK Parliament Web Archive	=12
US Library of Congress Web Archive	=14

Participants were also asked if there were any other web archives that they accessed or used, and to name the resource as free text. 1 user respondent indicated ‘Yes’ and provided the following free text: “Unsure whether scientific journal and search engine archives count here -I use them” (User). One might consider that this respondent has a hazy understanding of the differences between a digital archive, a digital library, and a web archive, and this could indicate that the respondent may be confused about being a web archive user, and indeed, may not be a user at all. However, the respondent identified using the Wayback Machine “to access older versions of websites. I'm interested in longitudinal data concerning water parameters, but sometimes websites only display current or recent results” (User). Thus, it would be fair to suggest that while an individual might use a web archive for their research, they may still have a blurred understanding of the differences between a digital archive, a digital library, and a web archive.

#### 4.4.6.5 Use of a dark web archive for studies/research

User respondents (n=59) were asked if they ever accessed or used a dark web archive for their studies/research. They were informed that a dark web archive is only accessible onsite in a designated reading room or Library via an onsite portal. Respondents were also asked to name the resources they used if they answered ‘Yes’. 94.92% (=56) of user respondents

indicated 'No', and 3 respondents indicated 'Yes' but did not provide any free text to name the dark web archive they used.

#### 4.4.6.6 Likelihood of future engagement by users with a dark web archive

Using a Likert scale, user respondents (n=59) were asked their opinions on the likelihood that they will access or use a dark web archive in the future for their studies/research. [Figure 4.5](#) provides an overview of responses. As one can see, 8.47% (=5) of user participants responded with 'Definitively Likely', 32.22% (=19) with 'Fairly Likely', and 11.86% (=7) with 'Unsure'. Compared with the non-user responses on the likelihood of using a dark web archive in the future ([Figure 4.3](#)), user respondents offer a more positive outlook for the likelihood of using a dark web archive in the future.

As mentioned previously, while the concept of a dark web archive, and its value as a research resource may not be clearly understood by many non-users, it also seems that some users may too have a hazy comprehension of the concept of a dark web archive, its value as a research resource, and how it might be used. Also, earlier findings revealed that only 2% (4 of 239) of the total number of respondents in the survey were aware of the existence of the NLI dark domain (.ie) web archive ([section 4.4.3](#)). Thus, there is a need for a collaborative effort in raising awareness of its existence and to foster discussions regarding its future access and use for Irish based research.

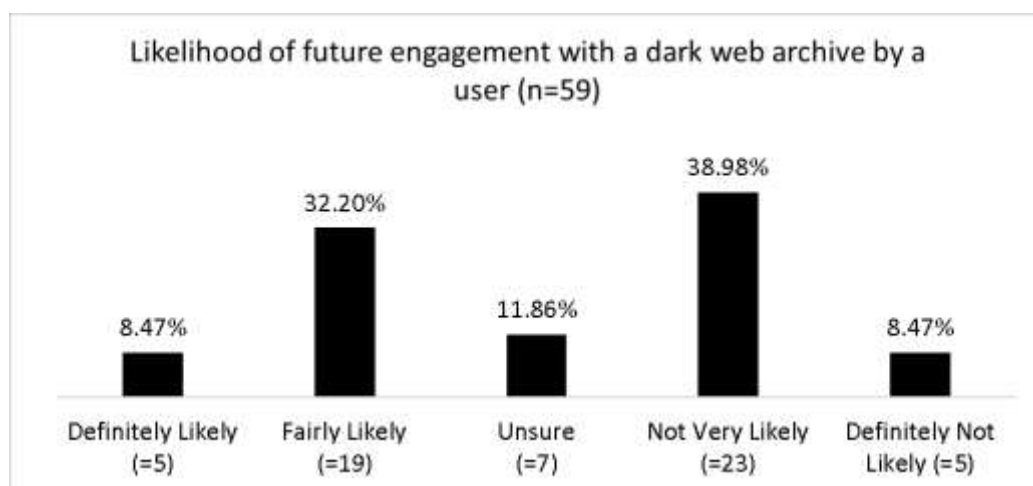


Figure 4.5: Representation for the likelihood of future engagement by users (n=59) with a dark web archive

#### 4.4.7 Perceived value and importance of web archives

This section provides an overview of total participant responses (N=239) to questions on the perceived value of web archives, the importance of archiving websites and blogs based on topics, and whether web archives will become important as a resource for current, medium, or long-term future research in their field.

##### 4.4.7.1 Perceived value of web archives

Using a Likert scale, and a list of six values, participants (N=239) were asked their opinion on the importance of archiving websites and blogs for current and future research, based on the values. Participants were offered a list of values with multiple choice options. [Table 4.18](#) provides a full breakdown of responses. Over half of all respondents indicated ‘Very Important’ for historical value (63.60%, =152), followed by research value (55.65%, =133), and evidential value (53.97%, =129); and just under half of respondents indicated ‘Very Important’ for cultural value (48.95%, =117).

Table 4.18: Representation of participant responses (N=239) for their perceived value of web archives

Participation per value (N=239)	Very important	Fairly important	Slightly Important	Not important	No opinion
Historical value	63.60% (=152)	22.59% (=54)	6.69% (=16)	1.67% (=4)	5.44% (=13)
Research value	55.65% (=133)	29.71% (=71)	5.86% (=14)	3.35% (=8)	5.44% (=13)
Evidential value	53.97% (=129)	25.94% (=62)	6.69% (=16)	3.77% (=9)	9.62% (=23)
Cultural value	48.95% (=117)	31.80% (=76)	10.88% (=26)	2.09% (=5)	6.28% (=15)
Technical value	25.52% (=61)	28.45% (=68)	25.10% (=60)	7.11% (=17)	13.81% (=33)
Design/artistic value	24.27% (=58)	24.69% (=59)	26.78% (=64)	11.30% (=27)	12.97% (=31)

##### 4.4.7.2 Perceived importance of archiving websites based on specific topics

Provided with a list of nine topics and a Likert scale, participants (N=239) were asked their opinion on the importance of archiving websites and blogs based on a topic area. [Table 4.19](#)

gives a breakdown of participant responses and shows that the most important topics to be archived are 'Direct Government' websites, deemed as 'Very important' (64.46%, =166) and 'Fairly Important' (19.25%, =46); and Indirect Government websites, deemed as 'Very important' (58.16%, =139) and 'Fairly Important' (28.87%, =69). Moreover, the archiving of Science and Environment websites/blogs were rated as more important than the archiving of websites on Referendums, Politics, Elections and Events. This could be used as an indicator for future collection development policies, as the inclusion of such topics may appeal to a wider academic audience and thus, attract a broader range of engagement. Moreover, it emphasises the need for a more rigorous approach for the inclusion of direct and indirect governmental websites as part of national digital heritage.

#### 4.4.7.3 Importance of web archives as a resource for current, medium, or long-term future research

Using a Likert scale, participants (N=239) were asked their opinion on whether web archives would become important as a resource for current, medium, or long-term future research in their field. [Table 4.20](#) offers a breakdown of participant responses.

Also, of interest are the discipline categories of respondents who indicated 'Yes', that web archives would become important.

- 98 respondents who indicated 'Yes' for current research identified with 20 discipline categories,
- 142 respondents who indicated 'Yes' for medium-term research identified with 20 discipline categories,
- 148 respondents who indicated 'Yes' for long-term research also identified with 20 discipline categories.

Table E.1 in [Appendix E](#) offers a full breakdown of discipline categories for the respondents mentioned above.

Table 4.20 clearly indicates that many participants feel that web archives will become more important for research as time goes on, and so, there is a need to establish theoretical and methodological approaches to enable researchers and educators to work with this type of data sooner rather than later. In addition, Appendix E demonstrates the need to consider potential research models and paradigms that are fit for purpose for a wide range of research fields.

Table 4.19: Representation of participant responses (N=239) on the importance of archiving websites/blogs based on topics

Participation per topic (N=239)	Very important	Fairly important	Slightly important	Not important	No Opinion
<b>Direct Government</b> (websites of official government departments or office holders, e.g., websites of the Department of Finance, the President, an Taoiseach)	69.46% (=166)	19.25% (=46)	4.18% (=10)	1.67% (=4)	5.44% (=13)
<b>Indirect Government</b> (websites of agencies deployed by the Irish Government to undertake a task, e.g., Irish Water, Nama)	58.16% (=139)	28.87% (=69)	5.86% (=14)	1.67% (=4)	5.44% (=13)
<b>Politics</b> (websites/blogs of political parties, political commentators)	50.63% (=121)	27.62% (=66)	10.88% (=26)	5.02% (=12)	5.86% (=14)
<b>Community Groups/ Activists</b> (websites/blogs of clubs, societies, advocacy groups, human rights groups)	42.26% (=101)	32.22% (=77)	13.39% (=32)	6.28% (=15)	5.86% (=14)
<b>Events</b> (websites/blogs for natural disasters, sporting events, commemoration events)	32.64% (=78)	38.49% (=92)	17.99% (=43)	4.60% (=11)	6.28% (=15)
<b>Election Campaigns</b> (websites/blogs of candidates, election judicators, commentators)	40.59% (=97)	30.54% (=73)	18.41% (=44)	5.02% (=12)	5.44% (=13)
<b>Referendum Campaigns</b> (websites/blogs of interest groups, referendum judicators, commentators)	46.86% (=112)	31.80% (=76)	12.55% (=30)	4.60% (=11)	4.18% (=10)
<b>Environment</b> (websites/blogs which report on climate change, pollution, conservation)	54.39% (=130)	31.38% (=75)	7.95% (=19)	2.93% (=7)	3.35% (=8)
<b>Science</b> (websites/blogs which report on advances in medicine, chemistry, physics)	57.32% (=137)	28.45% (=68)	6.28% (=15)	4.18% (=10)	3.77% (=9)

Table 4.20: Representation of participant responses (N=239) on the importance of web archives for current, medium, or long-term future research

Participant responses on the importance of web archives for current, medium, or long-term future research	Yes	Maybe	No
Current research (next 5 years) (N=239)	41.00% (=98)	36.40% (=87)	22.59% (=54)
Medium-term research (5-15 years) (N=239)	59.41% (=142)	29.71% (=71)	10.88% (=26)
Long-term research (15+ years) (N=239)	61.92% (=148)	27.62% (=66)	10.46% (=25)

#### 4.4.8 Perceived challenges for the use of archived web content for studies or research in the future

Finally, at the end of the survey, participants were provided with an optional open-ended question, and asked their opinion on their perceived challenges for the future use of archived web content in their field of research. 49 respondents entered free text of which 14 identified as a user, and 35 as a non-user. The free text was coded through the number of times a particular challenge was mentioned in participants' answers. For example, one participant may mention multiple challenges in one response, and thus, each individual challenge mentioned is included as a representation (R/r=).

Table 4.21 offers a breakdown of the thematic representations of responses by participants (n=50) on their perceived challenges for the future use of archived web content in their field of research and is organised into four main sub-themes as follows:

- Using web archives and archived web content (r=40)
- Awareness of the existence, content, and value of web archives (r=16)
- Data management and preservation (r=11)
- Not relevant for research topic (r=4)

These sub-themes are discussed in more detail in the next section.

Table 4.21: Thematic representation of participant responses on their perceived challenges for the future use of archived web content in their field of research (n=50)

Theme representation of responses for perceived challenges for the future use of archived web content (n=50)	No. of coded representations (R=71)
> Using web archives and archived web content <ul style="list-style-type: none"> <li>• Search and navigation (r=10)</li> <li>• Volume of data (r=10)</li> <li>• Access and discovery (r=9)</li> <li>• Representativeness and completeness of the data (r=5)</li> <li>• Non-established source/source credibility (r=4)</li> <li>• Citation (r=2)</li> </ul>	r=40
> Awareness <ul style="list-style-type: none"> <li>• Awareness of the existence, content, and value of web archives (r=16)</li> </ul>	r=16
> Data management and preservation <ul style="list-style-type: none"> <li>• Data management and data reliability (r=5)</li> <li>• Storage (r=3)</li> <li>• Technical challenges (r=3)</li> </ul>	r=11
> Not relevant for research topic	r=4

#### 4.4.8.1 Challenges Using web archives and archived web content

The responses presented several representations on the challenges for using web archives and archived web content in different contexts (r=40). These representations are further broken down into the sub-themes below.

- Search and navigation (r=10)
- Volume of data (r=10)
- Access and discovery (r=9)
- Representativeness and completeness of the data (r=5)
- Non-established source/source credibility (r=4)
- Citation (r=2)

##### Search and navigation

10 representations refer to challenges on how to use a web archive in terms of search and navigation. Some examples are outlined below.



- “Searching through the mass of material without a traditional kind of curated catalogue” (User, Social Sciences)
- “Individual researcher knowledge about how to use search functions on web archives” (Non-user, Natural Sciences)
- “searchability of archived data” (Non-user, Engineering Science)
- “Techniques for searching” (Non-user, Social Sciences)
- “navigation hopefully by content indexing” (User, Computer Science)
- “how to [...] search through it” (User, Social Sciences)
- “how to navigate them” (Non-user, Humanities)

### Volume of data

10 representations mention challenges in dealing with large volumes of data. Out of this, 3 representations refer to big data analytics, but from different perspectives. 1 representation notes a lack of training for humanities researchers using large scale data, while 2 representations show some concern about the use of big data analytics. Some examples of these representations are provided below.

- “Weeding the wheat from the chaff due to the sheer volume of information” (Non-user, Social Sciences)
- “The very large quantity of material” (Non-user, Humanities)
- “The volume of material available for a single organisation or an event (e.g. an election) may exceed the volume available for similar events or organisations in the past. Researchers will therefore have to deal with a far greater level of data than their predecessors” (User, Humanities)
- “The scale of the data available and the current deficit of training in Ireland in tools for large scale data analysis for humanities researchers.” (User, Humanities),
- “I would be concerned that the increasing use of data mining and other techniques to analyse large volumes of such content will result in only a partial analysis/understanding of any topic as the absence/balance of material may be misunderstood/misinterpreted.” (User, Dental Science)
- “The quantity of information available and the difficulty of understanding reception and audience. In some senses these are the same problems encountered by historians of the 20th C already in terms of print but they are on a very large scale. There may be a temptation to move towards increased emphasis on quantification thus losing some of the nuance and interest of qualitative approaches.” (Non-user, Humanities)

### Access and discovery

9 representations refer to accessibility, access, or discovery with 1 user respondent specifically mentioning access to the NLI domain web archive. Another representation

suggests that web archives will be underutilised if they are not findable through search engines, especially for those who are unaware of their existence. Some examples of these representations are outlined below.

- “Ensuring open access regardless of location e.g. if the National Library makes its [domain] web archive access [...] in a single location this would probably be in Dublin and non-Dublin based researchers would then have very limited access” (User, Engineering Science)
- “how to access them” (Non-user, Educational Science),
- “The accessibility [...] of archived data” (Non-user, Engineering Science),
- “Access and technological limitations” (Non-user, Computer Science)
- “Are these archives harvested by engines like Google? If not, they may be under-utilised as people who are not aware of them may not go directly to the archive to search for relevant material” (User, Humanities)
- “accessibility” (User, Humanities)
- “how to access them (Non-user, Educational Science)

#### Representativeness and completeness of the data:

There are 4 representations which mention the representativeness of the data in web archives, in terms of what is presented (or not presented) on the web and what ends up in a web archive. Examples are provided below.

- “My specific topic of research is controversial and scientific/ professional/commercial organisations tend to keep a minimum of information about it online. On the other hand, political/community/environmental groups with opposing views are vocal and prolific online. In my experience, research using archived web content will be constrained by what does and does not become web content, and these decisions and the reasons for them may not be evident from a simple examination of the content that is available. In other words, the context in which the content is created could be lost over time, even if the content itself is preserved.” (User, Dental Science)
- “The representativeness of the archived web content” (Non-user, Health Studies/Sciences)
- “Determining how representative the [...] content is for the general public or particular groups and their actual opinions, behaviors, and/or thoughts” (User, Business, Economics, Finance)

1 representation notes the completeness of the data due to chronological gaps in capture dates

- “In the case of the Wayback Machine, the unpredictable (and occasionally somewhat erratic) frequency of archiving can be problematic - you might have several snapshots in a month, and then miss all of the following year.” (User, Media/Communications)

#### Non-established source and source credibility

4 representations mention, in some way, that the use of web archives and archived web content is not an established or credible source for research. Some examples are provided below.

- “most students and researchers as far as I know tend to use more traditional sources (books, journals, patents) and since this is an established method of research it will be hard to break.” (Non-user, Engineering Science)
- “It is a largely unknown entity. It might be difficult to convince my supervisors that it is a valid source of credible information.” (Non-user, Social Sciences)
- “The credibility of such sources may be raised... Questions around why they are no longer active etc which likely have perfectly reasonable reasons but could cause doubt” (Non-user, Heritage Studies, Archival Studies)
- “The content of a web archive would not be a 'credible' citation in the legal field” (Non-user, Law)

#### Citation

2 representations note citation as a challenge for the use of archived web content.

- “researcher knowledge of how to cite archives and versions of historical web pages viewed on archives so that research is reproducible.” (Non-user, Natural Sciences)
- “Citation systems” (Non-user, Humanities)

#### 4.4.8.2 Awareness of the existence, content, and value of web archives

In terms of challenges for future engagement, there are 16 representations which refer to awareness in the context of awareness of the existence, the content, and the value of web archives for research. Several examples are outlined below.

- “How to increase awareness of their existence” (Non-user, Lecturer, Humanities)
- “public awareness of the archives” (User, Humanities)
- “Making people aware of such archives both within university/college information systems and in public libraries/information kiosks/public service web portals” (Non-user, Computer Science)
- “It is a largely unknown entity” (Non-user, Social Sciences)
- “How to increase awareness of their existence” (Non-user, Humanities)

- “Aware of the content” (Non-User, Educational Science)
- “Knowing that they exist” (Non-user, Architecture)
- “knowing where and what is there” (Non-user, Medicine, Biomedical Engineering)
- “Making researchers aware of relevant and informed content” (Non-user, Engineering Science)
- “Demonstrating value to future users as there is a habit now of searching the web for everything [...] You need to show people papers, projects where the value of an archive is demonstrable or at some type of culture night thing or something so people learn from such archives” (Non-User, Computer Science)

#### 4.4.8.3 Data management and preservation

5 representations mention data management, reliability, storage, or preservation in some context. Some examples are provided below.

- “proper and appropriate maintenance of the content” (Non-user, Professor or Associate Professor, Humanities)
- “Data management of these research resources. Loss of important documents like email communications will be a factor too re non-archiving of these tech files ” (Non-user, Researcher, Humanities)
- “Ensuring that the content has not been tampered with” (User, Humanities)
- “cost of storage” (Non-user, Built Environment)
- “Storage capacity” (User, Computer Science)

3 representations note technical challenges in terms of the capture and preservation of web content, and the provision of access to archived content in the context of the changing nature of browsers, platforms, and software. Examples are provided below.

- “I study and research social (digital) media and new apps and platforms are constantly being developed, while others close or become defunct. Archiving content on closed or defunct sites will be difficult if/when the technology advances to the point where these are no longer compatible with the latest operating systems, browsers, or mobile devices.” (User, Media/Communications)
- “technical access (risks associated with obsolescence of software or devices)” (Non-user, Humanities)
- “Access and technological limitations (e.g. browser support for older technology, etc)” (Non-user, Computer Science)

#### 4.4.8.4 Not relevant for research topic

4 representations infer that the use of archived web content in their field of study was due to the non-relevance of a web archive for their research. Some examples are provided below.

- “My field is not directly served by the above categories (ancient history).” (Non-user, Humanities)
- “Important data will remain online. There will be no need for ‘wasting’ time digging in old files - I am talking about science. For humanities, it might differ.” (Non-user, Natural Sciences)
- “I just don't think they're relevant to science. We access peer-reviewed scientific literature. Web archives have a place, but not in my research.” (Non-user, Natural Sciences)

### 4.5 Discussion

The survey results and analysis are based on a final number of 239 respondents, of which 59 respondents identified as a user (24.69%), and 180 respondents identified as a non-user (75.31%). The term user and non-user relates to whether a respondent specified that they have used or not used a web archive for their research or studies. The survey collected enough quantitative data combined with an element of qualitative data to provide useful insights for a discussion to address some of the research questions proposed for this study.

First, the findings shows that a large majority of all respondents (N=239) use the web for research, of which 88.28% of respondents indicated ‘Always’, and 11.30% indicated ‘Sometimes’, demonstrating that the web is a major research resource in Irish third-level academic institutions (see [Table 4.7](#)).

#### 4.5.1 Current level of awareness for the existence of web archives

Respondent (N=239) awareness of online public web archives differed, of which the Internet Archive, Wayback Machine was the most widely known resource (31.38%), followed by the US Library of Congress Web Archive (23.85%), and the public UK Web Archive (21.76%). 18.41% (n=44) of respondents indicated awareness of the public NLI Web Archive, and identified as educators, students, and researchers, from 11 different discipline categories.<sup>47</sup>

---

<sup>47</sup> \*Business, Economics, Finance; \*Computer Science; \*Digital Arts/Humanities/Heritage; \*Educational Science; \*Engineering Science; \*Geography; \*Humanities; \*Law; \*Media/Communications; \*Natural Sciences; \*Social Sciences

Awareness of the online public NLI Web Archive vis-à-vis the rate of respondents per discipline category was highest for the Humanities; however, it was quite low for other discipline categories such as the Social Sciences, Engineering Science, and the Natural Sciences.

While there is a large gap between the awareness of the Wayback Machine (31%), and both the NLI Web Archive (18%) and PRONI web archive (13.8%), Riley and Crookston (2015) also found a gap in their New Zealand study in so far as, much more respondents were aware of the Wayback Machine, than were aware of the New Zealand Web Archive (p. 12). Riley and Crookston (2015) submit that this may be due to the high profile of the Wayback Machine, poor efforts to promote the New Zealand Web Archive, and because their web archive collections are not in a standalone resource, rather they are integrated in a common interface which includes other library collections (p. 12).

In the case of the online NLI Web Archive and PRONI Web Archive, the gap may similarly be due to a lack of promotion but could also be due to a lack of use cases in Irish third-level education and research, which might showcase the use of the resources and thus, encourage more use. Moreover, the NLI only began a web archiving initiative in 2011, and PRONI in 2010, and as such, they are relatively young archives in comparison to some others. Therefore, it is encouraging to see some levels of awareness for these archives, which can be built upon for further promotion, outreach, and collaboration.

On the other hand, awareness of the NLI domain web archive is quite poor at 2%, and thus, will warrant a strategy for promotion as a research resource, when it eventually becomes accessible. In this regard it will be essential for the NLI to be afforded the capacity to collaborate with users and promote the resource to potential users; and to build solid research infrastructures between the NLI web archive, and the research teams seeking to use the data (Brügger, 2021c). This will require funding, and a cultural shift placing the creator and user as partners in the full web archiving lifecycle.

In addition, as noted by one respondent, access to a domain web archive onsite in the NLI reading room only, will present a geographical barrier for some researchers. Maurer (2022) also discusses how the provision of onsite 'only' access to web archive collections makes them geographically inaccessible for many researchers. Moreover, Healy et al. (2022) discuss how onsite access may present barriers for engagement due to socio-economic reasons (p. 114). Therefore, in terms of the establishment of a future Irish domain web archive, the obvious solution to the access problem would be to make it open access using an 'Opt-Out' strategy. However, this is probably unlikely for all types of web content. Therefore, for content that requires restrictions, such as content behind paywalls, there will be a need to consider how

access can be provided in more than one geographic location, perhaps in conjunction with other legal deposit libraries in Ireland. Moreover, access provisions should be made for researchers and users who are not affiliated to an academic institution. In the long-term, access should be provided in public libraries across Ireland, and this would ensure that users are not disadvantaged based on geographic location or socio-economic circumstances.

It must also be emphasised that certain categories of websites should be open access by default, including:

- (i) websites belonging to the Irish government, its departments, and its subsidiary agencies, as well as local government and councils,
- (ii) websites belonging to public bodies, quangos, civic agencies, and political parties who receive government funding in any form,
- (iii) websites belonging to owners or organisations who have received funding from the Irish government or any of its subsidiary agencies, and this should be stipulated as part of any funding agreement, and
- (iv) websites which have a variety of Creative Commons licences could also be considered for inclusion for open access.

#### 4.5.2 Terminology

As mentioned in [section 4.3.4](#), the data from 46 survey participants was not included in the final analysis of this survey study; nonetheless, it still warrants inclusion in the discussion regarding awareness of web archives as resources for research. There is ample evidence from the 46 survey respondents to suggest that they were confused as to what a web archive is, and for the most part, respondents correlated the meaning of a web archive to that of a digital library, digital archive, or digital data repository. Healy et al. (2022) also found a similar occurrence of this in the WARST survey (pp. 22–23), corresponding with the observations of Costea (2018) that the term web archive may not be “self-explanatory” enough for some researchers, and this could be due to “an ongoing lack of audience familiarity with the source” (p. 11). Brügger (2018) also highlights the difficulty with the term. He discusses whether a web archive best fits to the family of an archive or library, but notes that while they may be misleading, the terms web archive and web archiving were coined decades ago and so, part of the vernacular for this type of resource (pp. 77–78).

#### 4.5.3 Reasons for a lack of engagement with web archives for research

Of total respondents (N=239), 75.31% (n=180) acknowledged that they did not engage with web archives for their research or studies. The reasons for this are varied; however, a large

majority of non-user respondents (141 of 180) indicated that non-engagement was due to a lack of awareness of the availability of web archives as resources for research. Furthermore, the findings suggest that the value of web archives as a research resource is not clearly understood by an unfamiliar audience. For some respondents, there is a need for more efforts to demonstrate the importance of archiving the web and to promote the value of web archives for research. Therefore, it could be suggested that there is a need for the dissemination of use cases in Irish based research that will demonstrate the use of web archives as a research resource. These findings compare well with the findings from other user studies, whereby Jatowt et al (2008), Riley and Crookston (2015), and Costea (2018) also found a large lack of awareness of the existence of web archives by the participants in their studies. Winters (2017) also points to a lack of awareness as being one of the major reasons as to why web archives are not being more utilised.

Challenges in using a web archive, and web archive content, for research were also mentioned by non-user respondents whereby 81 participants did not know how to find archived websites relevant to their studies/research in a web archive, and 75 participants did not know how to use a web archive for their studies/research. While Costea's (2018) study concludes that a lack of scholarly use of web archives is related to a lack of awareness of their existence, Costea also notes that many researchers were unaware of the content of a web archive, and how a web archive can be used as a resource for research (p. 25). Thus, the findings of this study also correlate to the findings by Costea (2018).

In terms of other reasons for a lack of engagement, 47 participants were unsure of the credibility or authority of using archived websites, which demonstrates pedagogical issues. In addition, 32 participants did not know how to cite/reference an archived website which correlates well with the findings from Healy et al. (2022), whereby individuals from both the web archiving community and the scholarly user community experienced challenges for the citation of archived websites or derived datasets of archived web content (pp. 104–106). 24 participants indicated that they were unsure about copyright implications for using archived web content, and interestingly only 14 respondents felt that they did not have the technical skills to use a web archive for their studies/research which seems quite low. However, one should consider here that the respondents are non-users and therefore, have yet to discover the various types of technical skills and tools which are required for participation in web archive research as a user of web archives (Healy et al., 2022; Schmid et al., 2022).



#### 4.5.4 Likelihood of a non-user using a web archive for research, after becoming aware of its existence

There is good reason to believe that creating awareness of the availability of online public web archives increases the likelihood of researcher engagement. With the use of filters, we demonstrated how awareness increases the probable likelihood of research engagement by non-users (n=180), with percentages of 44.40% for the NLI Web Archive, 26.81% for the Wayback Machine, and 24.84% for the UK Web Archive. However, while the findings suggest that awareness increases the likelihood for an increase in researcher engagement, there is an indication that the promotion of the existence of web archives by itself may not be enough. For example, for some respondents, more efforts are needed to demonstrate the research value of web archives. Again, this highlights a need for more use cases in Irish based research to demonstrate approaches for using web archives as a resource, and the need for research infrastructures between web archive creators and web archive users to assist in promoting the value and use of these resources. In terms of the likelihood of a non-user respondent using a dark web archive for their research in the future, the response rate is low. However, it is also suggested that one needs to account for the fact that many of these respondents indicated that they did not use web archives for their research due to a lack of awareness of their existence (see [Table 4.10](#)). Thus, it could simply be a case that the concept of a dark (domain) web archive, and its value as a research resource is not clearly understood.

#### 4.5.5 Challenges perceived by scholars for the future use of archived web content

Of all participants (N=239), 50 respondents provided text responses for the challenges they perceived for future engagement with archived web content in their research fields. The text was analysed and broken down into four main themes and sub-themes as outlined below.

- Using web archives and archived web content
  - Search and navigation
  - Volume of data
  - Access and discovery
  - Representativeness and completeness of the data
  - Non-established source/source credibility
  - Citation
- Awareness of the existence, content, and value of web archives
- Data management and preservation
  - Data management and data reliability

- Storage
- Technical challenges
- Not relevant for research topic

The perceived challenges presented by respondents are certainly useful in understanding how we might proceed to incorporate the use of web archives for teaching and for conducting Irish based research, alongside more traditional sources, and methods. Of interest are the different outlooks on the use of large-scale analysis. 1 respondent notes the need for training in big data analysis for Humanities, while 2 respondents are concerned that big data analysis does not account for a full understanding of the context of the data. Rather, this might be better achieved with a qualitative approach. This certainly implies that there is a need to consider research models that consider both qualitative and quantitative methods as standalone practices, or a mixture of both as a combined approach to include web archives as a resource for research in Ireland.

#### 4.5.6 Users of web archives in Irish academic institutions

The survey results did not present any significant patterns to suggest that nationality, age, or gender has any influence on engagement with web archives. Of respondents who identified as a user (n=59), 30 respondents identified as educators, 19 as students and 10 as researchers. The data also shows that user respondents identified with 17 discipline categories.<sup>48</sup> As there has been a recent growth in the literature which promotes the use of web archives as resources for research in the humanities and social sciences (Gomes et al., 2021b; Brügger & Laursen, 2019; Brügger & Milligan, 2019; Brügger, 2018; Milligan, 2019; Brügger & Schroeder, 2017; Ogden, 2022; Gorsky, 2015), it is perhaps no surprise to see a strong number of users from the 'Humanities' in this study. On the other hand, there was a low-level of users from the 'Social Sciences'. The findings show that this is most likely due to a lack of awareness of the existence of web archives.

In terms of using a web archive in general, a large majority of user respondents (n=59) indicated that they use web archives for research interests (93.22%, =55). Respondents also indicated the use of a web archive for personal interests (72.88%, =43), historical interests (64.41%, =38) and cultural interests (42.37%, =25).

---

<sup>48</sup> Architecture; Business/Economics/Finance; Computer Science; Dental Science; Digital Arts/Humanities/Cultural Heritage; Educational Science; Engineering Science; Geography; Government/Public Administration; Humanities; Law; Mathematics; Media/Communications; Natural Sciences; Nursing/Midwifery; Political Science; Social Sciences

User respondent reasons for using archived web content for their studies or research are further outlined in [Table 4.22](#) and indicate that user respondents utilise web archives and archived web content for coursework purposes, for professional publication and historical research purposes, for qualitative and qualitative research purposes, for teaching purposes, and for access to materials no longer available on the live web. What is also surprising is that users come from a diverse range of research fields, and thus, this reflects that both multidisciplinary and interdisciplinary deliberation is required to consider the challenges, and potential solutions, for developing research models and paradigms for the use of web archives for Irish based research that are fit for purpose in a broad spectrum of research fields. Stember's (1991) description of the terms multidisciplinary and interdisciplinary is a useful guide here. For Stember (1991), multidisciplinary entails a collaboration between individuals from different disciplines "who each provide a different perspective on a problem or issue", and interdisciplinary is a step up from that to entail a collaboration between individuals from different disciplines to integrate methods and knowledge "into harmonious relationships" through a synthesis of strategies and approaches (p. 4).

Certainly, the user responses in this study offer some valuable insights on the opportunities for the use of web archives for Irish based research, and there is reason to believe that this community will grow over the next few years, as more academics become aware of web archives as resources for research. However, increases in web archive engagement will also depend on the promotion of awareness of the value of web archives, and demonstrations of use cases in academia as well as the public sphere. Formulating an Irish based multidisciplinary/interdisciplinary research network for current scholarly users and potential users, as well as web archivists, information professionals and technicians, and design professionals would be of great benefit here. It would assist in addressing potential solutions for developing research models and paradigms for the use of web archives for Irish based research that are fit for purpose in a broad spectrum of research fields. It would further enable discussions to develop frameworks to provide course modules for students in the use of web archives for research, and training courses for educators on how to incorporate web archived content as part of their teaching materials and methods.

Table 4.22: Combined data from Section 3.6 for user participant (n=59) reasons for using archived web content for their studies or research

<b>For coursework purposes</b>
<ul style="list-style-type: none"> <li>• as a primary source in an academic essay/assignment for my course</li> <li>• to document the history of an organisation in an academic essay/assignment for my course</li> <li>• as a secondary source for a thesis</li> </ul>
<b>For professional publication and historical research purposes</b>
<ul style="list-style-type: none"> <li>• as a primary source in a professional research report</li> <li>• as a primary source in a professional publication</li> <li>• to document the history of an organisation in a professional report/publication</li> </ul>
<b>For teaching purposes</b>
<ul style="list-style-type: none"> <li>• as part of teaching materials for undergraduate students</li> <li>• as part of teaching materials for postgraduate students</li> <li>• to recover old advertisements for teaching</li> </ul>
<b>For qualitative and quantitative research purposes</b>
<ul style="list-style-type: none"> <li>• for content analysis / textual analysis / discourse analysis</li> <li>• for data mining / topic modelling / data visualisation</li> <li>• for network analysis / geo-spatial analysis</li> </ul>
<b>For access to materials no longer available on the live web</b>
<ul style="list-style-type: none"> <li>• for accessing content / websites no longer available on the web</li> <li>• for access to “policy pages that were no longer publicly accessible, so as to compile evidence”</li> <li>• for a study of “longitudinal data concerning water parameters”</li> </ul>

#### 4.5.7 Perceived importance of archiving websites based on specific topics

Regarding the perceived importance of archiving websites based on a topic area, participants (N=239) considered Direct Government websites and Indirect Government to be of highest importance. Of further interest, respondents rated the archiving of Science websites and Environment websites as more important than websites on Politics, Referendums, Elections and Events. This could be used as an indicator for future collection development policies, as the inclusion of such topics may appeal to a wider academic audience and thus, attract a broader range of engagement. It also emphasises the need for a more rigorous approach for the inclusion of direct and indirect governmental websites as part of national digital heritage.

#### 4.5.8 Perceived value of web archives

Of total participant responses (N=239), 63.60% of respondents perceived the historical value of a web archive to be 'Very Important', followed by research value (55.65%), evidential value (53.97%) and cultural value (48.95%). Regarding whether web archives would become important as a resource for current, medium, or long-term future research in their field, many participants indicated that web archives will become more important for research as time goes on. So, there is a need to establish theoretical and methodological approaches, to enable researchers and educators to work with this type of data sooner rather than later. Furthermore, these participants identified with 20 disciplines, again demonstrating the need to consider potential research models and paradigms that are fit for purpose for a wide range of research fields.

#### 4.6 Summary

Overall, the findings demonstrated a limited awareness of the existence of web archives in Irish academic institutions, and, for an unfamiliar audience, more effort is needed to demonstrate the importance of archiving the web and to promote the value of web archives as resources for research. On the other hand, the findings revealed that there is a small community of web archive users in Irish academic institutions, at different levels of education and academia, aged from 18 to 65 years, and from a broad range of research fields. So, there is already a starting base of scholarly users and potential users which could be built upon to promulgate discourse for developing multidisciplinary and interdisciplinary research networks with web archivists, information professionals and technicians, as well as web design professionals to address potential solutions for developing research models and paradigms for the use of web archives for Irish based research that are fit for purpose in a broad spectrum of research fields. It would also enable discussions to develop frameworks to provide course modules for students in the use of web archives for research, and training courses for educators on how to incorporate web archived content as part of their teaching materials and methods.

## 5. CONCLUSIONS

This report is the product of a collaborative study by Sharon Healy (Maynooth University) and Helena Byrne (British Library). The study incorporated a review of resources and literature, informal dialogues with heritage colleagues, and the use of an online survey. The study sought to (i) examine the causes for the loss of digital heritage and how this relates to Ireland, (ii) offer an overview of the landscape of web archives based across Ireland, and their availability, and accessibility as resources for Irish based research, and (iii) provide some insight into the awareness of, and engagement with, web archives in Irish third-level academic institutions. In doing so, the report explored the relationship between legal deposit legislation and the preservation of national heritage and observed how web archiving is a necessary activity for the preservation of digital heritage. Furthermore, through a synthesis of legal deposit history, web history, and political debates, the report illustrates the need for the continual evaluation of legal deposit legislation in line with the fragility of born digital heritage and the technological advances in publishing and communication technologies.

While Brügger (2018) points to various ways in which the web might be archived, and while it is important to acknowledge such efforts, the focus of this study, for the most part, was on institutional web archiving and curation through web crawling, and the use of institutional web archives for research or other purposes. In addition, the study considered Irish digital heritage to be inclusive of the digital heritage of the island of Ireland, and when required, it referred to the digital heritage of Northern Ireland (NI) or the Republic of Ireland (ROI) to distinguish between the two jurisdictions.

In this final section, the report concludes with an overview of the key findings and insights, through an examination of the main causes for the loss of digital heritage, the availability and accessibility of web archives based on the island of Ireland for conducting Irish based research, and scholarly engagement and non-engagement with web archives in Irish academic institutions. The report concludes with some final thoughts for future work.

### 5.1 Key Findings and Insights

#### 5.1.1 Main causes for the loss of digital heritage

Within a few years of the web becoming established as a new medium for publishing and sharing information, national libraries and cultural heritage organisations became concerned about the ephemeral nature of the web, and instigated preservational strategies for the

capture and preservation of digital heritage on the web through web archiving. [Section 2.1](#) offered an overview of how these concerns were further substantiated by studies which examine link rot, reference rot, and web content change over time. There are several reasons put forward as to why web content moves, changes, or gets deleted, including software and system upgrading, changes in filing systems, the re-arrangement of web content, the re-location of servers, a lack of funding or interest to maintain websites, and simply a lack of foresight by web publishers. [Section 2.1](#) and [section 2.2](#) also demonstrated how concerns for the loss of digital heritage on the web stemmed from wider concerns about the appraisal, storage and long-term preservation of electronic information, multimedia and born digital materials in general. Moreover, these wider concerns have been around since before the web was invented, with the web just becoming another media carrier to worry about.

UNESCO (2003) posits that “the disappearance of heritage in whatever form constitutes an impoverishment of the heritage of all nations”, and digital heritage should not be an exception. Some of the factors which contribute to the loss of digital heritage to posterity include technological obsolescence of hardware and software, media deterioration, availability of resources, and inadequate legislation (UNESCO, 2003; Waters & Garrett, 1996; Besser, 2000). Moreover, the loss of digital heritage has often gone unnoticed by societies and nations because “Attitudinal change has fallen behind technological change” and consequently, the economic, social, intellectual, and cultural value or potential value of the heritage is not realised (UNESCO, 2003). For Lyman (2002), societies have lost important parts of their cultural heritage in the past because it was not archived or preserved due to cultural, technical, economic, and legal problems. The cultural problem is due to the inability of past generations to recognise its importance and historic value, while the technical problem is due to a lack of foresight and technical ingenuity to ensure continuity for preservation, storage, and maintenance (Lyman, 2002). Lyman (2002) posits how the economic problem stems from the failure to find a business model to support the archiving of new media formats, while the legal problem stems from the failure to create legislation which protects copyright while at the same time allows for archival preservation. These problems equally apply to the loss of digital heritage on the web. Although, it could be argued that the web archiving community has come a long way in providing solutions to the technical problem. Nonetheless, as internet and web technologies keep evolving, the capture tools will always be trying to catch up (Truman, 2016).

[Section 2.2](#) further demonstrated how the evolving nature of publishing over the past 50 years became problematic for legal deposit legislation which was fundamentally print-centric. For hundreds of years the concept of legal deposit served as a system to compile and preserve

a collection of a country's publications outputs, thus, providing a significant contribution to national cultural heritage. As a result, several countries began to amend their copyright and legal deposit legislation from the 1990s to accommodate the deposit of non-print materials and media formats (e.g., microfilm, CD-ROM, DVD etc.) and for born digital materials, inclusive of the web archiving of a country's national web domain, as a matter of routine. On the other hand, for many countries legal deposit legislation is still outdated in line with emerging publishing technologies and the advances in internet and web technologies. For instance, the ROI has trailed behind Canada, New Zealand, and much of Europe (Conul, 2012, p. 14). Also, in relation to the ROI, section 2 highlighted how there have been continual warnings by the National Archives of Ireland (NAI) to the ROI government, since at least 1997, regarding the loss of digital heritage due to the lack of a "comprehensive formal records, management policy for State" and the "Loss of electronic records and archives or access to them, due to degeneration of storage media and/or redundancy of operating systems" (Reports of the Director of the National Archives of Ireland, 2014-2020). Regrettably, over twenty years since the problem was identified, the Irish government has still not come to terms with the preservation of electronic records, nor does it seem to have a formal policy for record keeping in any electronic format. In addition, section 2.1 demonstrated how content on the Irish government website(s) has changed and disappeared over the past decades, while [section 3.4](#) established how Irish government department websites have been particularly vulnerable to link rot, and changes in website content.

Finally, section 3.4 underscored one of the major causes for the loss of Irish digital heritage to posterity, is the failure of successive ROI governments to negotiate copyright and legal deposit legislation in line with advances in publishing and communications technologies, exacerbated by the current deficiencies of ROI copyright and legal deposit legislation to include the routine web archiving of the Irish national domain as part of a national legal deposit scheme. In terms of the web space of NI, [section 3.3](#) illustrated how the UK legal deposit legislation was reformed in 2003, to allow for a selective web archiving initiative (undertaken by the UK Web Archiving Consortium) which also incorporated the capture and preservation of websites from the NI web space. Moreover, the legislation was updated again in 2013 to allow for an annual web crawl of the UK web estate (undertaken by the UK Web Archive), inclusive of the NI web space. It also demonstrated how the PRONI Web Archive commenced a selective web archiving initiative in 2010 to capture and preserve websites of NI government departments, local councils, public sector organisations and websites which have social, cultural, political, religious, or economic significance for the preservation of NI heritage. However, prior to 2013, the UK/NI web space was not systematically captured as part of legal deposit, and therefore much of the earlier NI webspace will have disappeared or



changed drastically (Jackson, 2015a). To salvage some of the UK web estate prior to 2013, the Joint Information Systems Committee (JISC) acquired a dataset from the Internet Archive which included all .uk websites in their web collection that were crawled from 1996-2013 (UK Web Archive, n.d., JISC UK Web Domain Dataset). The JISC UK Web Domain Dataset is available for use through the UK Web Archive website and listed in the British Library Shared Research Repository.

### 5.1.2 Availability and accessibility of web archives based on the island of Ireland for conducting Irish based research

Section 3 examined the availability and accessibility of web archiving initiatives based on the island of Ireland, and their usefulness for conducting Irish based research. In doing so, it offered insights which may be useful when it comes to assessing support and incentive mechanisms for scholarly researchers using web archives and other types of end users. While the section acknowledged that Irish web heritage can be found in various international web archives, the focus was on web archiving initiatives which have a specific mandate to capture a wide range of Irish web heritage as part of their collection development strategies. Therefore, the focus was on the PRONI Web Archive, the NLI Web Archive, and the UK Web Archive, which is accessible onsite in the Library of Trinity College Dublin (TCD). The section observed the efforts of these initiatives for the collection and preservation of digital heritage from the web spaces of NI and ROI, and offered an overview of their historical backgrounds, inclusive of how copyright and legal deposit has influenced their collecting activities.

In the case of the digital heritage of NI, the findings suggest that, while there are resource and legislative limitations, there are nonetheless concrete efforts being made to provide a balanced approach towards the collection and preservation of the NI web space. First, the UK Web Archive captures and preserves websites from the NI web space, through a selective collection approach and through an annual domain crawl of the NI web space as part of legal deposit, which is accessible onsite in a UK legal deposit library, inclusive of the Library of TCD in Dublin. Second, NI digital heritage is preserved through a two-fold approach by PRONI to provide a publicly accessible selective web archive collection, through (i) the collection of websites of government, public bodies etc., with notifications of the intent to collect, and provisions of a takedown policy, and (ii) a permissions-based approach for privately funded websites. And third, the NI web space is preserved through a collaborative effort by the PRONI Web Archive and the UK Web archive for the development of accessible curated collections. While there is a wide range of topics within the collections of the UK Web Archive and the PRONI Web Archive which would be useful for conducting Irish based research, access to the collections differ. PRONI Web Archive is open access, and the UK Web Archive is a mix of both

open access and onsite access. However, as discussed, onsite access presents challenges for researchers due to the restrictive nature of the access protocols in the current UK legal deposit legislation, which is outdated in line with advances in publishing and communications technologies, and current trends in digital user expectations and information seeking behaviours (Gooding et al., 2019).

Regarding the ROI, section 3 established how the National Library of Ireland (NLI) began a small-scale selective web archiving initiative in 2011, to include a wide range of topics which would be useful for conducting Irish based research. However, the section also highlighted how the NLI conducted two domain crawls in 2007 and 2017 which are currently inaccessible to researchers or the public due to legislative matters. While the NLI is a legal deposit library, digital legal deposit legislation was not enacted in Ireland at the time the domain crawls were conducted. Moreover, while digital legal deposit legislation came into force in December 2019 through the Copyright and Other Intellectual Property Law Provisions Act 2019 (hereafter, COIPLPA, 2019), it did not include a clause for crawling the Irish national web domain. However, COIPLPA (2019) does contain a clause to “bring forward a report on the feasibility of establishing a digital legal deposit scheme to serve as a web archive for .ie domain contents and advise on steps taken towards that goal” within twelve months of the Act coming into force in December 2019. However, as of March 2023, a feasibility report has yet to be produced.

The section further discussed how the establishment of a ROI national domain web archive is a necessary component for the preservation of Irish national digital heritage and examined some of the political debates regarding the inclusion of the archiving of the Irish national web domain as part of legal deposit legislation in line with other countries. It outlined that, while there are traces of the ROI web estate in other web archives, these are very shallow. Therefore, it will be impossible to retrospectively recreate the ROI web space. As it stands, the ROI is already “impoverished” (UNESCO, 2003) due to mass losses of digital heritage on the web for the decades of the 1990s, 2000s, and 2010s. It now looks like this will continue well into the 2020s, before the necessary measures are put in place for the collection and preservation of the web space of the twenty-six counties of the ROI in line with the collection and preservation of the web space of the six counties of NI. Therefore, it was stressed that urgent action is required for an emergency change in ROI legislation to allow for the collection and preservation of the ROI web estate in the interim, while a feasibility report continues to be undertaken to advise on the necessary requirements to update the legislation, and to establish a national web domain archive through “a process of negotiation among interested parties” (Lyman, 2002). Moreover, as demonstrated, negotiations should be inclusive of

representatives from the education and teaching sectors, end users who use web archives for wide range of purposes, information professionals who have experienced the transition from small-scale selective web archiving to large-scale domain web archiving, and information professionals who are experienced in working with Irish based information ecosystems. The section further emphasised the need to assess the demarcation of the Irish national web domain, as using the .ie ccTLD is not an adequate marker for the representation of Irish digital heritage on the web. Finally, the section underscored how born digital content is more fragile than print material, and publishing and communications technologies are constantly changing, thus legal deposit legislation needs to be reviewed on a regular basis in order to keep up with the changes in technology and current trends in digital user expectations and information seeking behaviours (Gooding et al., 2019). We further noted how the formation of a Copyright Council of Ireland, as suggested in the Modernising Copyright report (2013), could be tasked with monitoring legal deposit legislation in line with the fragility of born digital heritage and the technological advances in publishing and communication technologies.

### 5.1.3 Scholarly engagement and non-engagement with web archives

In the next sections we sum up the challenges and opportunities for using web archives as resources for research and provide an overview of the findings on scholarly awareness, engagement, and non-engagement with web archives in Irish academic institutions.

#### 5.1.3.1 Challenges for scholarly engagement with web archives

Section 2.3 put forward several reasons for the lack of scholarly engagement with web archives. Obvious reasons include a lack of awareness, or simply because some academic disciplines have no need to rely on such sources (Jatowt, 2008; Riley & Crookston, 2015; Winters, 2017; Costea, 2018). It can also be argued that a lack of dialogue or collaboration between the creators of web archives, and end users (or even potential end users) has had some effect on engagement with web archives for research purposes, as initially, web archiving initiatives tended not to prioritise how web archives would or might be used as part of their web archiving strategies (Dougherty et al., 2010; Hockx-Yu, 2014; Schroeder & Brügger, 2017; Gooding et al., 2021). Thus, Truman (2016) stresses the need for more communication and collaboration between those who curate, create and steward web archives and those who use (or might use) a web archive for purposeful research.

Challenges also arise, due to the characteristics of an archived website or web page which may not be a complete surrogate of what was once on the live web, rather, it is a version

(Brügger, 2010). Deficiencies in the archived artefacts may occur because of the temporal dimensions such as the time it takes to capture, and the possibility of content updates during capture. Deficiencies may also occur due to technical issues such as glitches during the archiving process such as robots.txt or limitations with the archiving software/hardware to keep up with the constant change and upgrade of web media file types and the evolving nature of dynamic content (Brügger, 2010; Meyer et al., 2011; Pennock, 2013; Maemura, 2018; Bingham & Byrne, 2021). In addition, in order to preserve a website or web page in its entire capacity to produce meaning, it should be inclusive of links to external (hyperlink) information, and quite often this is not achieved due to selection criteria, acquisition policies, technical glitches, financial constraints, or legislative and copyright restrictions (Besser, 2000; Milligan, 2019; Hockx-Yu, 2014). Finally, the collected web content may undergo technical processes during collection, preservation and to provide access through replay or playback (Brügger 2016, 2018; Schneider et al., 2009). This is why Brügger (2019; 2018; 2016) describes archived web content as reborn digital media, which is clearly distinct from other types of archived media such as film, television, photographs, and newspapers. Therefore, this implies that the use of archived web content for scholarly purposes has ongoing pedagogical challenges.

Other commentators note challenges due to the variances between searching on the live web, and searching in a web archive (Costa, 2021; Holzmann & Nejd, 2021; Winters & Prescott, 2019; Jackson et al., 2016b; Nielsen, 2016). The findings through web archive search techniques also tend to present multiple copies of content captured during different crawls, so they have a temporal dimension, which manifests more challenges. Both Brügger (2016) and Schafer (2019) suggest that web archives present challenges due to the “absence” of a traditional style catalogue or registry as an entry point. Costea (2018) identifies a need for improvements to web archives in the areas of discoverability options, data selection, data management, and access to more comprehensive documentation and metadata. Challenges for researchers/users also arise due to a lack of technical knowledge in the application of data mining techniques to vast volumes of data, as well as a lack of training and experience in using web archives from discovery processes to integrating the use of archived web content with traditional research approaches (Truman, 2016). Researchers wishing to take a more qualitative approach towards using the archived web, also have challenges due to a lack of research methods and theoretical paradigms for the use of the archived web (Millward, 2015). Other challenges relate to the fact that some large-scale web archives, such as the Internet Archive’s Wayback Machine, may lack depth and are deemed as too broad to meet the needs of specific research which often requires precise datasets (Schneider et al., 2009; Dougherty & van den Heuvel, 2009). Therefore, researchers often turn to developing their

own web archive collections for their needs (see for example, Foot & Schneider, 2006; Engholm, 2000). However, such collections are often narrow in scope and may never be useful for anything other than the study for which they were created (Dougherty & van den Heuvel, 2009).

Legislation on copyright and legal deposit also presents challenges for researchers to utilise web archives. Using the UK Web Archive legal deposit collections as an example, scholars discuss the challenges in using legal deposit collections which are only accessible on a library terminal in a designated reading room. Such challenges include the locked down nature of the library terminal whereby researchers cannot view the source code or copy the URL from the browser which causes problems for citation (Winters, 2020a; Milligan, 2015). Users are not allowed to copy and paste text which totally disrupts the affordances that are used by researchers worldwide, when they use the live web as a source for research (Milligan, 2015). Also, users can not take photographs or screenshots of the screen, rather they must pay for a printout of an archived web page, which is ironic, as researchers are allowed to use cameras to take photographs of historical documents in most archival environments (Milligan, 2015). Furthermore, no two people can view the same instance of an archived web page simultaneously which inhibits collaborative research as well as the use of the resource for teaching in the context of classroom group projects (Winters, 2020). Such challenges are manifested due to the restrictive nature of the UK legal deposit legislation as laid out in The Legal Deposit Libraries (Non-Print Works) Regulations 2013 (NPLD). Gooding et al. (2019) also discuss the challenges with the NPLD access protocols and highlight how the NPLD regulations make no allowance for text or data mining, and how this presents a barrier for innovative research. Furthermore, Gooding et al. (2021) suggest that the user was neglected as a stakeholder when it came to drafting the legislation for NPLD access protocols, which is fundamentally print-centric. Moreover, they insist that because the NPLD ethos is print-centric, it fails to consider the user in line with digital user expectations, and current trends in information seeking behaviours (Gooding et al., 2021). Therefore, when it comes to evaluating resources like legal deposit collections, in particular the use of collections with restrictions, it needs to be clearly examined in relation to the rapidity in which technology changes the landscape for end users.

There are other implications regarding the use of web archives with access restrictions. Maurer (2022) and Healy et al. (2022) note how the provision of onsite 'only' access to web archive collections in a designated building makes web archives geographically and socio-economically inaccessible for many researchers. Furthermore, Truter (2021) highlights the challenges for end user researchers in terms of the access and use of archived web content

due to legal restrictions, inclusive of copyright and third-party ownership, privacy policies, and the General Data Protection Regulation (GDPR) in the European Union (EU). This manifests challenges for not only the use of the data, but also affects how and if the data can be made shareable and reusable (Truter, 2021) and runs counter to the requirement of open science which is being stipulated by a growing number of research institutions and funding agencies (Winters, 2020a).

Challenges for researchers also arise due to ethical, sociotechnical, and political circumstances. Maemura (2018) points to challenges due to “ethical implications of how materials are used”, as well as “questions of consent” and the responsibility of the researcher to the people represented in the data. Ogden et al. (2022) suggest that researchers need to be vigilant using web archives when researching socially vulnerable communities, and Mackinnon (2021) warns researchers of the ethical implications when it comes to the study of websites of “young people of the past” and their right to be forgotten. Ogden and Maemura (2021) examine how the sociotechnical, organisational, and resource constraints “under which most web archiving programmes operate” needs to be understood by researchers, and suggest that researchers need to become familiar with the “specific limits and constraints, legal governance frameworks, collection mandates, as well as configurations (i.e. of sub-collections) and terminology used for specific collections.” In terms of political circumstances, Ben-David (2019) discusses the challenges for studying web histories of countries that do not have a ccTLD, such as Kosovo, which was denied the allocation of a ccTLD as it was not recognised as a sovereign state by the United Nations, due to a Russian veto.

Researchers may also be more interested in using big data methods such as topic modelling or network analysis on a web sphere of websites (WARC files) from a specific web archive collection (e.g., Geocities) or to do a longitudinal study across multiple legal deposit annual web domain collections (see Milligan, 2019; Brügger et al., 2017; Brügger et al., 2019). However, Maurer (2022) points out that organising large volumes of WARC files for research is difficult for both web archiving initiatives, and end user researchers. Reasons for this are varied and may be “due to a mix of curatorial, technical, legal, economic and organisational constraints” (Brügger, 2021c). Brügger (2021c) further stresses the need for solid research infrastructures between the web archives with the data, and the research teams wishing to use the data, to help overcome some of the legal, ethical, and technical challenges for both communities. This will require funding, and a cultural shift placing the creator and user as partners, in the full web archiving lifecycle.

One should also consider how some of the challenges mentioned above overlap between creators and users. For example, Healy et al. (2022) suggest that both creators and users have challenges in the areas of search and retrievability, users find it difficult to search large-scale web archives, while creators find it difficult to provide search mechanisms and algorithms for large-scale collections that will satisfy a diversity of users. Moreover, both creators and users have challenges with legal issues such as copyright and legal deposit, users have challenges accessing content, while creators have challenges for the collection of web content, as well as the provision of access, and how restrictive this access might be (Healy et al., 2022).

### 5.1.3.2 Scholarly use of web archives

Section 2.3 argued that a lack of dialogue or collaboration between the creators of web archives, and end users (or even potential end users) has had some effect on engagement with web archives for research purposes. On a more positive note, section 2.3 also discussed how collaboration between web archive creators and end user researchers has been improving over the past decade (Schroeder & Brügger, 2017; Webster 2017b; Maemura, 2022). This is partly due to growing efforts to foster and increase research engagement by consortiums, networks, research projects and libraries in some instances. Thus, collaborations between those who create the data (Brügger, 2021c) and those who want to use the data is proving to be a proactive solution for increasing scholarly engagement with web archives. Section 2.3 also demonstrated the growing number of journal publications, conference papers, and conference presentations which discuss the use of web archives as resources for research, or which offer case studies in the use of web archives and archived content. These cover topics such as media and journalism, social sciences and ethnographies, public health and telemedicine, information science and law, internet studies, web histories and more. Also of benefit, section 2.3 described an applicable model for studying the archived web using Brügger's (2018) five strata analytical grid. For Brügger (2018), the web can be examined through an analytical grid of five strata: an individual web element, an individual web page, an individual website, a web sphere, the web in its entirety (p. 31). Moreover, studies of the web may have an overlap of strata cases (Brügger, 2018, p. 68). For Brügger (2018) the five strata can be applied equally to both layers of the web being "the visible/audible web in the browser, and hidden text of HTML code and associated files" (p. 31). These five strata also offer an equally applicable model for studying the archived web.

Section 4 observed how literature integrating the use of archived web content for Irish based topics or research is difficult to find, with some exceptions being Malone (n.d.), Harjani (2018), Byrne (2019), Greene & Ryan (2019), Healy (2019), Webster (2019), and Greene (2020). Hence, it is also a useful starting point when considering the type of research that has

already been undertaken using web archives for research on Irish based topics, and how it can be built upon. It demonstrates the use of a qualitative approach (Malone, n.d., Healy, 2019), a big data approach (Greene & Ryan, 2019; Greene, 2020) and combining qualitative and big data approaches (Harjani, 2018; Byrne, 2019; Webster, 2019). This provides a good indicator on the types of research which needs to be accounted for in any forthcoming legislation on copyright and legal deposit in the ROI. Moreover, Harjani's (2018) research highlights the important role that social media can play in understanding key events in Irish society. Thus, any new legal deposit legislation introduced in the ROI should consider making provisions for the inclusion of social media content.

### 5.1.3.3 Scholarly awareness, engagement, and non-engagement with web archives in Irish academic institutions

Section 4 offered some insights into the challenges and solutions for using web archives for research in an Irish context. Through an online survey of lecturers, researchers, and students in Irish academic institutions, the section set out to provide some insight into the awareness of, and engagement with, web archives in Irish third-level academic institutions, in a bid to gain a better understanding of how and why archived web content is used or not used for research in Ireland. The section was also exploratory in terms of assessing some of the opportunities and challenges for using web archives, and considerations for how to best facilitate their use, going forward. Most prominently, the findings demonstrated a limited awareness of the existence of web archives in Irish academic institutions, and that creating awareness increases the probable likelihood for an increase in researcher engagement. However, the findings suggested that promoting awareness of the existence of web archives by itself may not be sufficient to impact engagement. For an unfamiliar audience, efforts are also needed to demonstrate the importance of archiving the web, the value of web archives for research, and more effort for awareness on how to use web archives for research. The findings also indicated that web archives will become more important for research as time goes on.

The survey findings also present several indicators on the challenges that scholars based in Ireland perceive for the future use of web archives and archived web content. How to use web archives and archived web content was presented as a challenge from several outlooks such as search and navigation, handling large volumes of data, citation practices for using archived web content, and research models for using web archives as a non-established source. Of interest are the different outlooks on the use of large-scale analysis, and the need for training in big data analysis for Humanities, while there are also concerns that big data analysis does not account for a full understanding of the context of the data. Rather, this



might be better achieved with a qualitative approach. This implies that there is a need to consider research models that consider both qualitative and quantitative methods as standalone practices, or a mixture of both as a combined approach to include web archives as a resource for research in Ireland. The completeness of the data was also mentioned in terms of capture frequencies, as well as challenges with the representativeness of the data in a web archive, in relation to what is presented (or not presented) on the web and what ends up in a web archive. There is also the case that data in a web archive is simply not relevant for a particular research discipline.

On a bright note, the findings also show that there is already a small community of web archive users in Irish academic institutions, aged from 18 to 65 years, and at different levels of education and academia. The findings further indicated that user respondents utilise web archives and archived web content for coursework purposes, for professional publication and historical research purposes, for teaching purposes, for qualitative and quantitative research purposes and for access to materials no longer available on the live web. What is also surprising is that users come from a diverse range of research fields, which reflects the need for both multidisciplinary and interdisciplinary deliberation to consider the challenges, and potential solutions, for developing research models and paradigms for the use of web archives for Irish based research that are fit for purpose in a broad spectrum of research fields.

Certainly, the user responses in this study offer some valuable insights on the opportunities for the use of web archives for Irish based research, and there is reason to believe that this community will grow over the next few years, as more academics become aware of web archives as resources for research. However, increases in web archive engagement will also depend on the promotion of awareness of the value of web archives, and demonstrations of use cases in academia as well as the public sphere. Formulating an Irish based multidisciplinary/interdisciplinary research network to comprise of current scholarly users and potential users, web archivists, information professionals and web design professionals would be of great benefit here. It would assist in addressing potential solutions for developing research models and paradigms for the use of web archives for Irish based research across a broad spectrum of research fields; and enable discussions to develop frameworks to provide course modules for students in the use of web archives for research, and training courses for educators on how to incorporate web archived content as part of their teaching materials and methods.

To end here, the survey findings indicated that awareness of the NLI domain archive is quite poor, and thus, will warrant a strategy for promotion as a research resource, when it

eventually becomes accessible. In this regard it will be essential for the NLI to be afforded the capacity to collaborate with users and promote the resource to potential users, and the capacity to build solid research infrastructures between the NLI web archive, and the research teams seeking to use the data (Brügger, 2021c). This will require funding, and a cultural shift placing the creator and user as partners in the full web archiving lifecycle. In addition, access to an Irish domain web archive onsite in the NLI reading room ‘only’ will present geographical and socio-economic barriers for some researchers (Maurer, 2022; Healy et al., 2022). Therefore, in terms of the establishment of an Irish domain web archive, the obvious solution to the access problem would be to make it open access using an ‘Opt-Out’ strategy. However, this is probably unlikely for all types of web content. Therefore, for content that requires restrictions, such as content behind paywalls, there will be a need to consider how access can be provided in more than one geographic location, perhaps in conjunction with other legal deposit libraries across Ireland. Moreover, access provisions should be made for researchers and users who are not affiliated to an academic institution. In the long-term, access should be provided in public libraries across Ireland, and this would ensure that users are not disadvantaged based on geographic location or socio-economic circumstances.

It must also be emphasised that certain categories of websites should be open access by default, including:

- (i) websites belonging to the Irish government, its departments, and its subsidiary agencies, as well as local government and councils,
- (ii) websites belonging to public bodies, quangos, civic agencies, and political parties who receive government funding in any form,
- (iii) websites belonging to owners or organisations who have received funding from the Irish government or any of its subsidiary agencies, and this should be stipulated as part of any funding agreement, and
- (iv) websites which have a variety of Creative Commons licences could also be considered for inclusion for open access.

## 5.2 Final Thoughts and Future Work

It is hoped that this report will serve as a starting point for fostering open dialogues across Ireland on the necessity for long-term preservation strategies for electronic information, multimedia, and born digital materials in general, with the web just being another media carrier to worry about. Moreover, in laying the groundwork, the report will contribute to the current debates regarding the necessity for the implementation of legal deposit legislation which realistically reflects the fragility of born digital heritage and the technological advances in publishing and communication technologies. The report also provides a starting point in

addressing some of the challenges regarding digital and web historiography and how this relates to Irish based research. As new methodologies are born out of necessity to deal with the advances of the internet, web and software technologies and the continual evolution of digital media, older methodologies will be doomed due to software incompatibilities or obsolescence and outdated digital media formats. This is not something new. Archivists, librarians, and information professionals have been discussing it for years. The big question here is how this affects the use of digital materials for academic research, whether they are digitised, born digital on the live web or reborn digital in a web archive, and how can we ensure the use of such digital materials remain accessible, allowing for research reproducibility in the future.

The Bibliography is organised in three parts: a list of **Primary Sources**, a list of **References**, and a list of web archiving **Providers & Services** mentioned throughout the report. The full Bibliography is also available in a Zotero web library.<sup>49</sup> We have tried to ensure that the URLs provided in the Bibliography and footnotes are (i) captured in a web archive close to the time of access on the live web or (ii) saved in a web archive close to the time of access on the live web. In case of future link rot, we have documented which web archive the URL may be found in, e.g., [URL Memento: Wayback Machine].

### Primary Sources

#### Acts, Bills, Amendments, Directives, Statutes

European Communities. Council Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society. European Parliament, 167 OJ L (2001). Retrieved 2022-10-31, from <http://data.europa.eu/eli/dir/2001/29/oj/eng> [URL Memento: Archive.today] EUR-Lex Doc ID: 32001L0029

European Economic Community. Council Directive 91/250/EEC of 14 May 1991 on the legal protection of computer programs. European Parliament, 122 OJ (1991). Retrieved 2022-09-28, from <https://op.europa.eu/en/publication-detail/-/publication/92d68447-ea9a-4554-9540-de517984c310/language-en>. [URL Memento: Wayback Machine]

European Economic Community. Council Directive 93/98/EEC of 29 October 1993 harmonizing the term of protection of copyright and certain related rights. European Parliament, 290 OJ L (1993). Retrieved 2022-09-28, from <http://data.europa.eu/eli/dir/1993/98/oj/eng>. [URL Memento: Wayback Machine]

Ireland. Copyright Act, 1963. No. 10/1963, Government of Ireland, Oireachtas (1963). Commencement Date: 01 October 1964. Retrieved 2022-02-06, from <https://www.irishstatutebook.ie/eli/1963/act/10/enacted/en/html>. [URL Memento: Wayback Machine]

Ireland. Eighth Amendment of the Constitution Act, 1983. No. C8/1983, Government of Ireland, Oireachtas (1983), Commencement Date: 07 October 1983. Retrieved 2021-09-23, from <https://www.irishstatutebook.ie/eli/1983/ca/8/enacted/en/index.html>. [URL Memento: Wayback Machine]

---

<sup>49</sup> Healy, S. & Byrne, H. (2022). Zotero Groups - Scholarly Use of Web Archives in Ireland. Zotero, [https://www.zotero.org/groups/4870080/scholarly\\_use\\_of\\_web\\_archives\\_in\\_ireland](https://www.zotero.org/groups/4870080/scholarly_use_of_web_archives_in_ireland)

- Ireland. National Archives Act, 1986. No. 11/1986, Government of Ireland, Oireachtas (1986). Commencement Date: 01 June 1988. Retrieved 2022-03-22, from <https://www.irishstatutebook.ie/eli/1986/act/11/enacted/en/html?q=National+Archives+Act> [URL Memento: Wayback Machine]
- Ireland. Dublin City University Act, 1989. No.15/1989, Government of Ireland, Oireachtas (1989). Commencement Date: 22 June 1989. Retrieved 2021-04-20, from <https://www.irishstatutebook.ie/eli/1989/act/15/enacted/en/html>. [URL Memento: Wayback Machine]
- Ireland. University of Limerick Act, 1989. No.14/1989, Government of Ireland, Oireachtas (1989). Commencement: 22 June 1989. Retrieved 2021-04-20, from <https://www.irishstatutebook.ie/eli/1989/act/14/enacted/en/html>. [URL Memento: Wayback Machine]
- Ireland. European Communities (Legal Protection of Computer Programs) Regulations, 1993. S.I. No. 26/1993 Government of Ireland, Oireachtas (1992). Commencement Date: 31 December 1992. Retrieved 2021-05-08, from <https://www.irishstatutebook.ie/eli/1993/si/26/made/en/print>. [URL Memento: Wayback Machine]
- Ireland. European Communities (Term of Protection of Copyright) Regulations, 1995. S.I. No. 158/1995, Government of Ireland, Oireachtas (1992). Commencement Date: 01 July 1995. Retrieved 2021-10-16, from <https://www.irishstatutebook.ie/eli/1995/si/158/made/en/print>. [URL Memento: Wayback Machine]
- Ireland. National Cultural Institutions Act, 1997, No.11/1997, Government of Ireland, Oireachtas (1997). Commencement Date: 02 June 1997. Retrieved 2022-01-19, from <https://www.irishstatutebook.ie/eli/1997/act/11/enacted/en/html>. [URL Memento: Wayback Machine]
- Ireland. Universities Act, 1997. No.254/1997, Government of Ireland, Oireachtas (1997). Commencement Date: 16 June 1997. Retrieved 2021-08-01, from <https://www.irishstatutebook.ie/eli/1997/act/24/enacted/en/index.html>. [URL Memento: Wayback Machine]
- Ireland. Copyright and Related Rights Act, 2000. No. 28/2000, Government of Ireland, Oireachtas (2000). Commencement Date: 01 January 2001. Retrieved 2019-03-27, from <https://www.irishstatutebook.ie/eli/2000/act/28/enacted/en/html>. [URL Memento: Wayback Machine]
- Ireland. Copyright and Other Intellectual Property Law Provisions Bill 2018 (as initiated). No. 31/2018, Government of Ireland, Oireachtas (2018). As initiated: 09 March 2018. Retrieved 2022-01-20, from <https://data.oireachtas.ie/ie/oireachtas/bill/2018/31/eng/initiated/b3118d.pdf> [URL Memento: Wayback Machine]
- Ireland. Thirty-sixth Amendment of the Constitution Act 2018. Act C36/2018 Government of Ireland, Oireachtas (2018), Commencement Date: 18 September 2018. Retrieved 2021-02-05, from <https://www.irishstatutebook.ie/eli/2018/ca/36/enacted/en/html>. [URL Memento: Wayback Machine]

- Ireland. Copyright and Other Intellectual Property Law Provisions Act 2019. No. 19/2019, Government of Ireland, Oireachtas (2019). Commencement Date: 02 December 2019 Retrieved 2021-10-26, from <https://www.irishstatutebook.ie/eli/2019/act/19/enacted/en/index.html> [URL Memento: Wayback Machine]
- Irish Free State. Industrial and Commercial Property (Protection) Act, 1927. No. 16/1927 Irish Free State, Oireachtas (1927). Commencement Date: 01 October 1927. Retrieved 2021-04-25, from <https://www.irishstatutebook.ie/eli/1927/act/16/enacted/en/html>. [URL Memento: Wayback Machine]
- Seanad Éireann. Copyright and Other Intellectual Property Law Provisions Bill 2018—Committee Stage—Amendments. No. 31b/2018, 03 October 2018, Houses of the Oireachtas, Ireland (2018). Retrieved 2022-01-20, from <https://data.oireachtas.ie/ie/oireachtas/bill/2018/31/seanad/3/amendment/numbered/List/eng/b31b18d-scnl.pdf> [URL Memento: Wayback Machine]
- Seanad Éireann. Copyright and Other Intellectual Property Law Provisions Bill 2018—Report Amendments. No. 31b/2018, 15 May 2019, Houses of the Oireachtas, Ireland (2019). Retrieved 2021-04-19, from <https://data.oireachtas.ie/ie/oireachtas/bill/2018/31/seanad/4/amendment/numbered/List/eng/b31b18d-srnl.pdf> [URL Memento: Wayback Machine]
- United Kingdom. Act of Union (Ireland) 1800. Acts of the Old Irish Parliament 1800 c. 38 (Regnal. 40\_Geo\_3), UK Parliament (1801). Commencement Date: 01 January 1801. <https://www.legislation.gov.uk/aip/Geo3/40/38/contents>. [URL Memento: Wayback Machine]
- United Kingdom. Public Records Act (Northern Ireland) 1923. Acts of the Northern Ireland Parliament 1923, Chapter 20, UK Parliament (1923). Commencement Date: 22 June 1923. Retrieved 2020-08-05, from <https://www.legislation.gov.uk/apni/1923/20/contents>. [URL Memento: Wayback Machine]
- United Kingdom. Legal Deposit Libraries Act 2003. UK Public General Acts 2003 c. 28, UK Parliament (2003). Commencement Date: 01 February 2004. Retrieved 2022-01-08, from <https://www.legislation.gov.uk/ukpga/2003/28/contents>. [URL Memento: Wayback Machine]
- United Kingdom. The Legal Deposit Libraries (Non-Print Works) Regulations 2013. UK Statutory Instruments 2013 No. 777, UK Parliament (2013). Commencement Date: 05 April 2013. Retrieved 2022-02-03, from <https://www.legislation.gov.uk/uksi/2013/777/contents/made>. [URL Memento: Wayback Machine]

## E-Mail

- IIPC curators list. (2017, April 4). Call for Participation: Archives Unleashed 4.0: Web Archive Datathon [distribution list communication].

## Newspapers

- Crowe, C. (2012, June 30). Ruin of Public Record Office marked loss of great archive. *The Irish Times* [online] Retrieved from <https://www.irishtimes.com/opinion/ruin-of-public-record-office-marked-loss-of-great-archive-1.1069843> [URL Memento: Wayback Machine]
- Cunningham, M. (1997a, March 17). Beware the ideas of March The Government Web site (<http://www.irlgov.ie>). *The Irish Times*, (COMPUTIMES), [CITY EDITION], p. 10. Proquest ID: 310308514
- Cunningham, M. (1997b, January 27). Brewster's millions. *The Irish Times*, (COMPUTIMES), [CITY EDITION], p. 18. ProQuest ID: 310187147
- Cunningham, M. (1997b, January 27). Brewster's millions. *The Irish Times*, (COMPUTIMES) [online] Retrieved from <https://web.archive.org/web/19990117002422/http://www.irish-times.com/irish-times/paper/1997/0127/cmp1.html>. [Wayback Machine, timestamp: 1999-01-17 00:24:22; URL source: <http://www.irish-times.com/irish-times/paper/1997/0127/cmp1.html>]
- Cunningham, M. (1995, January 2). 1994: Year of the Net Michael Cunningham chronicles some of the main computing stories during annus mulllmedius. *The Irish Times*, (COMPUTIMES), [CITY EDITION], p. 17. ProQuest ID: 309949150
- Fagan, K. (2012, May 21). When the server becomes the master. *The Irish Times* [online] Retrieved 2022-10-31, from <https://www.irishtimes.com/business/sectors/when-the-server-becomes-the-master-1.522745>. [URL Memento: Wayback Machine]
- McGee, H. (2018, December 28). Move to begin in 2019 to release State papers after 20 years. *The Irish Times*. Retrieved 2018-12-28, from <https://www.irishtimes.com/news/politics/move-to-begin-in-2019-to-release-state-papers-after-20-years-1.3742288> . [URL Memento: Wayback Machine]
- O'Connell, H. (2022, March 13). Sinn Féin wipes years of media statements from website. *Sunday Independent* [online]. Retrieved 2022-03-13, from <https://www.independent.ie/irish-news/news/sinn-fein-wipes-years-of-media-statements-from-website-41440873.html>. [URL Memento: Wayback Machine]
- Shenton, H. (2020, April 29). Digital black hole in our national memory. *The Irish Times* [online]. Retrieved 2020-05-14, from <https://www.irishtimes.com/opinion/letters/digital-black-hole-in-our-national-memory-1.4240247> [URL Memento: Wayback Machine]
- Ramesh, R., & Hern, A. (2013, November 13). Conservative party deletes archive of speeches from internet. *The Guardian* [online]. Retrieved 2020-06-13, from <https://www.theguardian.com/politics/2013/nov/13/conservative-party-archive-speeches-internet>. [URL Memento: Wayback Machine]
- Taylor, C. (2017a, July 20). Ireland's digital content in danger of disappearing, specialist warns. *The Irish Times* [online]. Retrieved 2017-09-22, from <https://www.irishtimes.com/business/technology/ireland-s-digital-content-in-danger-of-disappearing-specialist-warns-1.3157792>. [URL Memento: Wayback Machine]

Weiss, R. (2003, November 24). On the Web, Research Work Proves Ephemeral. *Washington Post* [online]. Retrieved 2018-04-29, from <https://www.washingtonpost.com/archive/politics/2003/11/24/on-the-web-research-work-proves-ephemeral/959c882f-9ad0-4b36-88cd-fb7411db118d/> [URL Memento: Wayback Machine]

## Irish Parliamentary Debates

Dáil Éireann. (1928). Ceisteanna—Questions. Oral Answers – Manuscripts Commission. – Dáil Éireann (6th Dáil) – Wednesday, 17 Oct 1928, Vol. 26 No. 4. Houses of the Oireachtas, Ireland. Retrieved 2022-10-24, from <https://www.oireachtas.ie/en/debates/debate/dail/1928-10-17/3>. [URL Memento: Wayback Machine]

Dáil Éireann. (2019). Departmental Websites – Dáil Éireann Debate – Tuesday 26 February 2019, Questions (7, 8, 9). Houses of the Oireachtas, Ireland. Retrieved 2019-10-24, from <https://www.oireachtas.ie/en/debates/question/2019-02-26/9/>. [URL Memento: Wayback Machine]

Dáil Éireann. (2021). Digital Archiving - Dáil Éireann Debate – Thursday 9 September 2021, Questions (267). Houses of the Oireachtas, Ireland. Retrieved 2022-09-14, from <https://www.oireachtas.ie/en/debates/question/2021-09-09/267>. [URL Memento: Wayback Machine]

Dáil Éireann. (2021). Intellectual Property – Dáil Éireann Debate – Thursday, 11 Nov 2021, Questions (175). Houses of the Oireachtas, Ireland. Retrieved 2022-09-29, from <https://www.oireachtas.ie/en/debates/question/2021-11-11/175>. [URL Memento]

Joint Committee on Tourism, Culture, Arts, Sport and Media. (2021a). Engagement with Chairperson Designate of the Board of the National Library of Ireland – Joint Committee on Tourism, Culture, Arts, Sport and Media debate – Wednesday, 13 Oct 2021 (Ireland). Houses of the Oireachtas, Ireland. Retrieved 2022-01-18, from [https://www.oireachtas.ie/en/debates/debate/joint\\_committee\\_on\\_tourism\\_culture\\_arts\\_sport\\_and\\_media/2021-10-13/2/](https://www.oireachtas.ie/en/debates/debate/joint_committee_on_tourism_culture_arts_sport_and_media/2021-10-13/2/). [URL Memento: Archive.today]

Joint Committee on Tourism, Culture, Arts, Sport and Media. (2021b). Key Priorities and Legislation of the Department of Tourism, Culture, Arts, Gaeltacht, Sport and Media: Discussion – Joint Committee on Tourism, Culture, Arts, Sport and Media debate – Wednesday, 24 Nov 2021 (Ireland). Houses of the Oireachtas, Ireland. Retrieved 2022-04-26, from [https://www.oireachtas.ie/en/debates/debate/joint\\_committee\\_on\\_tourism\\_culture\\_arts\\_sport\\_and\\_media/2021-11-24/3](https://www.oireachtas.ie/en/debates/debate/joint_committee_on_tourism_culture_arts_sport_and_media/2021-11-24/3). [URL Memento: Wayback Machine]

Joint Committee on Tourism, Culture, Arts, Sport and Media. (2022). Future of Media Commission Report: Discussion Joint – Joint Committee on Tourism, Culture, Arts, Sport and Media debate – Wednesday, 14 Sep 2022. Retrieved 2022-09-29, from [https://www.oireachtas.ie/en/debates/debate/joint\\_committee\\_on\\_tourism\\_culture\\_arts\\_sport\\_and\\_media/2022-09-14/2](https://www.oireachtas.ie/en/debates/debate/joint_committee_on_tourism_culture_arts_sport_and_media/2022-09-14/2). [URL Memento: Wayback Machine]



- Seanad Éireann. (2018). Copyright and Other Intellectual Property Law Provisions Bill 2018: Committee Stage – Wednesday, 3 Oct 2018, Vol. 260 No. 6. Houses of the Oireachtas. Retrieved 2021-12-24, from <https://www.oireachtas.ie/en/debates/debate/seanad/2018-10-03/14> [URL Memento: Wayback Machine]
- Seanad Éireann. (2021). An tOrd Gnó - Order of Business – Seanad Éireann Debate – Thursday, 30 Sep 2021, Vol. 278 No. 9. Houses of the Oireachtas, Ireland. Retrieved 2021-12-15, from <https://www.oireachtas.ie/en/debates/debate/seanad/2021-09-30/8>. [URL Memento: Wayback Machine]
- Seanad Éireann. (2021). Nithe i dtosach suíonna - Commencement Matters – Digital Archiving - Seanad Éireann Debate – Tuesday, 23 Nov 2021, Vol. 280 No. 7. Houses of the Oireachtas, Ireland. Retrieved 2021-11-29, from <https://www.oireachtas.ie/en/debates/debate/seanad/2021-11-23/3>. [URL Memento: Wayback Machine]

## Screenshots

- Beyond 2022. (n.d.). Virtual Tour: Record Treasury of Ireland. Beyond 2022. Retrieved 2022-10-31, from <https://vrtour.virtualtreasury.ie/> [URL Memento: Archive.today]
- Bragg, M., & Hanna, K. (2013). The Web Archiving Lifecycle Model [White Paper]. The Archive-It Team, Internet Archive. Retrieved 2021-10-07, from [http://ait.blog.archive.org/files/2014/04/archiveit\\_life\\_cycle\\_model.pdf](http://ait.blog.archive.org/files/2014/04/archiveit_life_cycle_model.pdf). [URL Memento: Wayback Machine]
- Department of Business, Enterprise and Innovation. (2017, September 5). Department of Business, Enterprise and Innovation – Home [Archived web page]. Department of Business, Enterprise and Innovation. Retrieved from <https://web.archive.org/web/20170905050758/https://dbei.gov.ie/en/#> [Web archive: Wayback Machine; source URL: <https://dbei.gov.ie/en/#>; Timestamp: 2017-09-05 05:07:58]
- Department of Enterprise, Trade and Employment. (1999, October 4). Department of Enterprise, Trade and Employment – Home [Archived web page]. Department of Enterprise, Trade and Employment. Retrieved from <https://web.archive.org/web/19991004144844/http://www.entemp.ie/> [Web archive: Wayback Machine; source URL: <http://www.entemp.ie/>; Timestamp: 1999-10-04 14:48:44]
- Department of Enterprise, Trade and Employment. (2002, May 29). Department of Enterprise, Trade and Employment – Home [Archived web page]. Department of Enterprise, Trade and Employment. Retrieved from <https://web.archive.org/web/20020529044054/http://www.entemp.ie:80/> [Web archive: Wayback Machine; source URL: <http://www.entemp.ie:80/>; Timestamp: 2002-05-29 04:40:54]
- Department of Enterprise, Trade and Innovation. (2010, May 7). Department of Enterprise, Trade and Innovation – Home [Archived web page]. Department of Enterprise, Trade

and Innovation. Retrieved from  
<https://web.archive.org/web/20100507101638/http://www.deti.ie/> [Web archive: Wayback Machine; source URL: <http://www.deti.ie/>; Timestamp: 2010-05-07 10:16:38]

Department of Jobs, Enterprise and Innovation. (2011, June 6). Department of Jobs, Enterprise and Innovation – Home [Archived web page]. Department of Jobs, Enterprise and Innovation. Retrieved from  
<https://web.archive.org/web/20110606063430/http://www.djei.ie/> [Web archive: Wayback Machine; source URL: <http://www.djei.ie/>; Timestamp: 2011-06-06 06:34:30]

Department of Jobs, Enterprise and Innovation. (2012, June 10). Department of Jobs, Enterprise and Innovation – Home [Archived web page]. Department of Jobs, Enterprise and Innovation. Retrieved from  
<https://web.archive.org/web/20120610184500/http://enterprise.gov.ie/>. [Web archive: Wayback Machine; source: <http://enterprise.gov.ie/>; Timestamp: 2012-06-10 18:45:00]

Department of Jobs, Enterprise and Innovation. (2012, June 13). Submissions Received 2012 on foot of the Copyright Review Consultation Paper [Archived web page]. Department of Jobs, Enterprise and Innovation. Retrieved from [https://wayback.archive-it.org/org-1444/20120613230622/http://www.djei.ie/science/ipr/crc\\_submissions2.htm](https://wayback.archive-it.org/org-1444/20120613230622/http://www.djei.ie/science/ipr/crc_submissions2.htm) [Web archive: Wayback Machine; source: URL: Timestamp: 2012-06-13 23:06:22]

Government of Ireland. (1996, December 24). Government of Ireland—Home [Archived web page]. Government of Ireland. Retrieved from  
<https://web.archive.org/web/19961224221829/http://www.irlgov.ie/> [Web archive: Wayback Machine; source URL: <http://www.irlgov.ie/>; Timestamp: 1996-12-24 22:18:29]

Government of Ireland. (1997, January 5). Government of Ireland – Home [Archived web page]. Government of Ireland. Retrieved from  
<https://web.archive.org/web/19970105013806/http://www.irlgov.ie/> [Web archive: Wayback Machine; source: URL: <http://www.irlgov.ie/>; Timestamp: 1997-01-05 01:38:06]

Government of Ireland. (2000, March 2). Government of Ireland – Home [Archived web page]. Government of Ireland - Home. Retrieved from  
<https://web.archive.org/web/20000302062006/http://www.irlgov.ie/> [Web archive: Wayback Machine; source: URL: <http://www.irlgov.ie/>; Timestamp: 2000-03-02 06:20:06]

Government of Ireland. (2002, March 28). Government of Ireland – Home [Archived web page]. Government of Ireland. Retrieved from  
<https://web.archive.org/web/20020328101202/http://www.irlgov.ie/> [Web archive: Wayback Machine; source: URL: <http://www.irlgov.ie/>; Timestamp: 2002-03-28 10:12:02]

Government of Ireland. (2008a, November 8). Government of Ireland – Home [Archived web page]. Government of Ireland. Retrieved from  
<https://web.archive.org/web/20081108001743/http://www.irlgov.ie/> [Web archive: Wayback Machine; source: URL: <http://www.irlgov.ie/>; Timestamp: 2008-11-08 00:17:43]

- Government of Ireland. (2008b, December 17). Government of Ireland – Home [Archived web page]. <https://web.archive.org/web/20081217033949/http://www.gov.ie/en/> [Web archive: Wayback Machine; source: URL: <http://www.gov.ie/en/>; Timestamp: 2008-12-17 03:39:49]
- Government of Ireland. (2011, July 3). Government of Ireland – Home. Retrieved from <https://web.archive.org/web/20110703223040/http://www.gov.ie:80/en/> [Web archive: Wayback Machine; source: URL: <http://www.gov.ie:80/en/>; Timestamp: 2011-07-03 22:30:40]
- International Internet Preservation Consortium. (2004, June 3). International Internet Preservation Consortium—Welcome (2004) [Archived web page]. International Internet Preservation Consortium. Retrieved from <https://web.archive.org/web/20040603014115/http://netpreserve.org/about/index.php> [Web archive: Wayback Machine; source URL: <http://netpreserve.org/about/index.php>; Timestamp: 2004-06-03 01:41:15]
- National Library of Ireland. (n.d.a). NLI Web Archive—Archive-It—Collections [Web page]. Archive-It. Retrieved 2022-07-04, from <https://archive-it.org/home/nli/?show=Collections> [URL Memento: Wayback Machine]
- National Library of Ireland. (n.d.b). NLI Web Archive—Archive-It—Sites [Web page]. Archive-It. Retrieved 2022-09-22, from <https://archive-it.org/home/nli/?show=Sites> [URL Memento: Wayback Machine]
- Public Record Office of Northern Ireland (PRONI). (n.d). PRONI Web Archive [Web page]. NI Direct Government Services. Retrieved 2022-10-21, from <https://webarchive.proni.gov.uk/#/> [URL Memento: Wayback Machine]
- UK Web Archive. (n.d.). UKWA Topics and Themes [Web page]. UK Web Archive. Retrieved 2022-06-19, from <https://www.webarchive.org.uk/en/ukwa/category/> [URL Memento: Wayback Machine]
- UNESCO & IFLA-PAC. (n.d.). Safeguarding our documentary heritage [Web page]. Webworld.Unesco.Org. Retrieved from <https://web.archive.org/web/20010429033355/http://webworld.unesco.org/safeguarding/en/index.html> [Web archive: Wayback Machine; source: URL: <http://webworld.unesco.org/safeguarding/en/index.html>; Timestamp: 2001-04-29 03:33:55]

## References

- Aasman, S. (2019). Finding Traces in YouTube’s Living Archive: Exploring Informal Archival Practices. *TMG Journal for Media History*, 22(1), 35–55. DOI: 10.18146/tmg.435
- Aasman, S., Haan, T. de, & Teszelszky, K. (2019). Web Archaeology: An Introduction. *TMG Journal for Media History*, 22(1), 1–5. DOI: 10.18146/tmg.433 [URL Memento: Wayback Machine]
- Adelmann, B., & Franken, L. (2020). Thematic web crawling and scraping as a way to form focussed web archives [Conference abstract]. *Engaging with Web Archives:*

- 'Opportunities, Challenges and Potentialities', (#EWAVirtual), Maynooth University Arts and Humanities Institute, Co. Kildare, Ireland, [online], 21-22 September 2022. Retrieved 2020-10-09, from <https://zenodo.org/record/4058013>. DOI: 10.5281/zenodo.4058013 [URL Memento: Wayback Machine]*
- Alberts, G., Went, M., & Jansma, R. (2017). Archaeology of the Amsterdam digital city; why digital data are dynamic and should be treated accordingly. *Internet Histories: Digital Technology, Culture and Society*, 1(1–2), 146–159. DOI: 10.1080/24701475.2017.1309852
- American Library Association. (2008, February 21). Definitions of Digital Preservation [Web page]. Association for Library Collections & Technical Services (ALCTS). Retrieved 2019-02-06, from <https://www.ala.org/alcts/resources/preserv/defdigpres0408>. [URL Memento: Wayback Machine]
- Analytical Access to the Domain Dark Archive. (2012+). Analytical Access to the Domain Dark Archive (AADDA) [Blog site]. Analytical Access to the Domain Dark Archive. Retrieved 2019-09-23, from <http://domaindarkarchive.blogspot.com>. [URL Memento: Wayback Machine]
- Anthony, A., Onasoga, K., Ike, D., & Ajayi, O. (2013). Web Archiving: Techniques, Challenges, and Solutions. *International Journal of Management & Information Technology*, 5(3), 598–603. DOI: 10.24297/ijmit.v5i3.760
- Antracoli, A., Duckworth, S., Silva, J., & Yarmey, K. (2014). Capture All the URLs: First Steps in Web Archiving. *Pennsylvania Libraries: Research & Practice*, 2(2), 155–170. DOI: 10.5195/palrap.2014.67 [URL Memento: Wayback Machine]
- Archive-It. (n.d.). Archive-It Blog [Blog site]. Archive-It. Retrieved 2021-09-01, from <http://ait.blog.archive.org/>. [URL Memento: Wayback Machine]
- Arvidson, A., Persson, K., & Mannerheim, J. (2000, August). The Kulturarw3 Project—The Royal Swedish Web Archiw3e—An example of 'complete' collection of web pages. *66th IFLA Council and General Conference Jerusalem, Israel, 13-18 August 2000*. IFLA Council and General Conference. Retrieved 2021-05-14, from <https://archive.ifla.org/IV/ifla66/papers/154-157e.htm>. [URL Memento: Wayback Machine]
- Aturban, M. (2019a, September 10). Where did the archive go? Part 2: National Library of Ireland [Blog post]. Web Science and Digital Libraries Research Group. Retrieved 2019-11-13, from <https://ws-dl.blogspot.com/2019/09/2019-09-10-where-did-archive-go-part-2.html>. [URL Memento: Wayback Machine]
- Aturban, M. (2019b, September 25). Where did the archive go? Part 3: Public Record Office of Northern Ireland [Blog post]. Web Science and Digital Libraries Research Group. Retrieved 2021-06-22, from <https://ws-dl.blogspot.com/2019/09/2019-09-25-where-did-archive-go-part-3.html>. [URL Memento: Wayback Machine]
- Aust, R. (2014, December 3). Online reactions to institutional crises: BBC Online and the aftermath of Jimmy Savile. In Publications Department [Paper]. Retrieved 2021-05-15, from <https://sas-space.sas.ac.uk/6100/>. [URL Memento: Wayback Machine]
- Bailey, J., Grotke, A., Hanna, K., Hartman, C., McCain, E., Moffatt, C., & Taylor, N. (2014). *Web Archiving in the United States: A 2013 Survey* [NDSA Report]. USA: National Digital

- Stewardship Alliance. Retrieved 2022-01-27, from <https://osf.io/h4e6z/>. [URL Memento: Archive.today]
- Bailey, J., Grotke, A., McCain, E., Moffatt, C., & Taylor, N. (2017). *Web Archiving in the United States: A 2016 Survey* [NDSA Report]. USA: National Digital Stewardship Alliance. Retrieved 2022-01-27, from <https://osf.io/hj7rg/>. [URL Memento: Archive.today]
- Bansal, S., & Parmar, S. (2020). Decay of URLs Citation: A Case Study of Current Science. *Library Philosophy and Practice*, 01–09. Retrieved 2020-10-30, from <https://digitalcommons.unl.edu/libphilprac/3582/>. [URL Memento: Wayback Machine]
- Bauer, R. (2018, August 2). What is the Difference Between Data Backup and Data Archive? [Blog post]. Backblaze Blog. Retrieved 2019-07-10, from <https://www.backblaze.com/blog/data-backup-vs-archive/>. [URL Memento: Wayback Machine]
- Beal, V. (2010). The Difference Between The Internet And World Wide Web. Webopedia (Online/web). Webopedia. Retrieved 2022-02-11, from <https://www.webopedia.com/insights/web-vs-internet/>. [URL Memento: Wayback Machine]
- Bearman, D. (1993). The Implications of ‘Armstrong v. Executive of the President’ for the Archival Management of Electronic Records. *The American Archivist*, 56(4), 674–689. JSTOR ID: 40293774
- Beaudouin, V., Pehlivan, Z., & Stirling, P. (2019). Exploring the Memory of the First World War Using Web Archives: Web Graphs Seen from Different Angles. In N. Brügger & I. Milligan (Eds.), *The SAGE Handbook of Web History* (pp. 440–463). London: SAGE Publications.
- Becker, C. (1938). What is Historiography? *The American Historical Review*, 44(1), 20–28. DOI: 10.2307/1840848
- Ben-David, A. (2019). National web histories at the fringe of the web: Palestine, Kosovo, and the quest for online self-determination. In N. Brügger & D. Laursen (Eds.), *The Historical Web and Digital Humanities: The Case of National Web Domains* ( pp. 89–109). London & New York: Routledge.
- Ben-David, A. (2021). Critical Web Archive Research. In D. Gomes, E. Demidova, J. Winters, & T. Risse (Eds.), *The Past Web: Exploring Web Archives* (pp. 181–188). Cham, Switzerland: Springer.
- Bergman, M. K. (2001). White Paper: The Deep Web: Surfacing Hidden Value. *Journal of Electronic Publishing*, 7(1). DOI: 10.3998/3336451.0007.104 [URL Memento: Wayback Machine]
- Berners-Lee, T. (1998). Hypertext Style: Cool URIs don’t change. [Web page]. W3C. Retrieved 2019-03-08, from <https://www.w3.org/Provider/Style/URI.html>. [URL Memento: Wayback Machine]
- Besser, H. (2000). Digital Longevity. In Maxine K. Sitts (Ed.), *Handbook for Digital Projects: A Management Tool for Preservation and Access*. Andover, Massachusetts: Northeast Big UK Domain Data for the Arts and Humanities. (n.d.).
- Beyond 2022. (n.d.). Beyond 2022—Ireland’s Virtual Record Treasury [Website]. Beyond 2022. Retrieved 2022-10-31, from [https://beyond2022.ie/?page\\_id=2](https://beyond2022.ie/?page_id=2) [URL Memento: Archive.today]

- Beyond 2022. (n.d.). Support [Web page] Beyond 2022. Retrieved 2022-10-31, from <https://www.virtualtreasury.ie/support> [URL Memento: Archive.today]
- Beyond 2022. (n.d.). Virtual Tour: Record Treasury of Ireland. Beyond 2022. Retrieved 2022-10-31, from <https://vrtour.virtualtreasury.ie/> [URL Memento: Archive.today]
- Big UK Domain Data for the Arts and Humanities [Website]. Big UK Domain Data for the Arts and Humanities. Retrieved 2021-04-19, from <https://buddah.projects.history.ac.uk/>. [URL Memento: Wayback Machine]
- Bingham, N. (2014). Quality Assurance Paradigms in Web Archiving Pre and Post Legal Deposit. *Alexandria: The Journal of National and International Library and Information Issues*, 25(1-2), 51–68. DOI: 10.7227/ALX.0020
- Bingham, N., Byrne, H., Lelkes-Rarugal, C., & Rossi, G. C. (2020, May 29). Using Webrecorder to archive UK political party leaders' social media after the UK General Election 2019 [Blog post]. UK Web Archive Blog. Retrieved 2021-06-19, from <https://blogs.bl.uk/webarchive/2020/05/using-webrecorder-to-archive-uk-political-party-leaders-social-media-after-the-uk-general-election-2.html>. [URL Memento: Wayback Machine]
- Bingham, N. J., & Byrne, H. (2021). Archival strategies for contemporary collecting in a world of big data: Challenges and opportunities with curating the UK web archive. *Big Data & Society*, 8(1), 1–6. DOI: 10.1177/2053951721990409
- Bødker, H., & Brügger, N. (2018). The shifting temporalities of online news: The Guardian's website from 1996 to 2015. *Journalism*, 19(1), 56–74. DOI: 10.1177/1464884916689153
- Boston, G. (Ed.). (1998). Memory of the World: Safeguarding the documentary heritage, a guide to standards, recommended practices and reference literature related to the preservation of documents of all kinds. General Information Programme and UNISIST United Nations Educational, Scientific and Cultural Organization. Retrieved from <https://unesdoc.unesco.org/ark:/48223/pf0000112676>. Document code: CII.98/WS/4
- Bragg, M., & Hanna, K. (2013). The Web Archiving Lifecycle Model [White Paper]. The Archive-It Team, Internet Archive. Retrieved 2021-10-07, from [http://ait.blog.archive.org/files/2014/04/archiveit\\_life\\_cycle\\_model.pdf](http://ait.blog.archive.org/files/2014/04/archiveit_life_cycle_model.pdf). [URL Memento: Wayback Machine]
- Breed, M. (2019). Capturing a Moment: The Practices and Ethics of Social Media Archiving [MA Thesis, The University of North Carolina at Chapel Hill University Libraries]. DOI: 10.17615/P4FF-ZK64
- Brown, A. (2006). Archiving Websites: A Practical Guide for Information Management Professionals. London: Facet Publishing.
- Brügger, N. (2010). Introduction: Web History, an Emerging Field of Study. In N. Brügger (Ed.), *Web History* (pp. 01–26). New York: Peter Lang.
- Brügger, N. (Ed.). (2010). *Web History*. New York: Peter Lang.
- Brügger, N. (2012). When the Present Web is Later the Past: Web Historiography, Digital History, and Internet Studies. *Historical Social Research / Historische Sozialforschung*, 37(4 (142)), 102–117. JSTOR ID: 41756477

- Brügger, N. (2016). Digital Humanities in the 21st Century: Digital Material as a Driving Force. *Digital Humanities Quarterly*, 10(2). Retrieved 2018-11-09, from <http://www.digitalhumanities.org/dhq/vol/10/3/000256/000256.html>. [URL Memento: Wayback Machine]
- Brügger, N. (Ed.). (2017). *Web 25: Histories from the first 25 Years of the World Wide Web* (Vol. 112). New York: Peter Lang.
- Brügger, N. (2018). *The Archived Web: Doing History in the Digital Age*. Massachusetts, London: The MIT Press.
- Brügger, N. (2019). Understanding the Archived Web as Historical Source. In N. Brügger & I. Milligan (Eds.), *The SAGE Handbook of Web History* (pp. 16–29). London: SAGE Publications.
- Brügger, N. (2020). Welcome to WARCnet. *WARCnet Papers*. Aarhus, Denmark: WARCnet. Retrieved 2021-01-27, from [https://cc.au.dk/fileadmin/user\\_upload/WARCnet/1.Bru\\_gger\\_Welcome\\_to\\_WARCnet.pdf](https://cc.au.dk/fileadmin/user_upload/WARCnet/1.Bru_gger_Welcome_to_WARCnet.pdf) [URL Memento: Wayback Machine]
- Brügger, N. (2021a). Digital humanities and web archives: Possible new paths for combining datasets. *International Journal of Digital Humanities*, 2(1), 145–168. DOI: 10.1007/s42803-021-00038-z
- Brügger, N. (2021b). Digital humanities and web archives: Possible new paths for combining datasets [Appendix]. OSF. Retrieved 2021-05-12, from <https://doi.org/10.17605/OSF.IO/ZSU3D>. [URL Memento: Wayback Machine]
- Brügger, N. (2021c). The Need for Research Infrastructures for the Study of Web Archives. In D. Gomes, E. Demidova, J. Winters, & T. Risse (Eds.), *The Past Web: Exploring Web Archives* (pp. 217–224). Cham, Switzerland: Springer.
- Brügger, N. (2021d). The WARCnet network: The first year. *WARCnet Papers*. Aarhus, Denmark: WARCnet. Retrieved 2021-08-20, from [https://cc.au.dk/fileadmin/user\\_upload/WARCnet/Bru\\_gger\\_The\\_first\\_year.pdf](https://cc.au.dk/fileadmin/user_upload/WARCnet/Bru_gger_The_first_year.pdf). [URL Memento: Wayback Machine]
- Brügger, N. (2022). Tracing a historical development of conspiracy theory networks on the web: The hyperlink network of vaccine hesitancy on the Danish web 2006–2015. *Convergence*, 13548565221104988. DOI: 10.1177/13548565221104989
- Brügger, N., & Finnemann, N. O. (2013). The Web and Digital Humanities: Theoretical and Methodological Concerns. *Journal of Broadcasting & Electronic Media*, 57(1), 66–80. DOI: 10.1080/08838151.2012.761699
- Brügger, N., & Laursen, D. (Eds.). (2019). *The Historical Web and Digital Humanities: The Case of National Web Domains*. London & New York: Routledge.
- Brügger, N., Laursen, D., & Nielsen, J. (2017). Exploring the domain names of the Danish web. In N. Brügger & R. Schroeder (Eds.), *The Web as History: Using Web Archives to Understand the Past and the Present* (Online/pdf, pp. 1–22). London: UCL Press. DOI: 10.14324/111.9781911307563 [URL Memento: Wayback Machine]
- Brügger, N., Laursen, D., & Nielsen, J. (2019). Establishing a corpus of the archived web. In N. Brügger & D. Laursen (Eds.), *The Historical Web and Digital Humanities: The Case of National Web Domains* (pp. 124–142). London & New York: Routledge.

- Brügger, N., & Milligan, I. (Eds.). (2019). *The SAGE Handbook of Web History*. London: SAGE Publications.
- Brügger, N., & Schroeder, R. (Eds.). (2017). *The Web as History: Using Web Archives to Understand the Past and the Present*. London: UCL Press. DOI: 10.14324/111.9781911307563; Online/pdf. [URL Memento: Wayback Machine]
- Burner, M., & Kahle, B. (1996, September 15). Arc File Format [Web page]. Internet Archive. Retrieved 2021-12-18, from <https://archive.org/web/researcher/ArcFileFormat.php>. [URL Memento: Wayback Machine]
- Byrne, H. (2019). Where are we now? A review of research on the history of women's soccer in Ireland. *Sport in History*, 39(2), 166–186. DOI: 10.1080/17460263.2019.1604422
- Byrne, H. (2020, September 10). Launching the UK Web Archive 2020 Annual Domain Crawl [Blog post]. UK Web Archive Blog. Retrieved 2021-05-30, from <https://blogs.bl.uk/webarchive/2020/09/launching-the-uk-web-archive-2020-annus-al-domain-crawl.html>. [URL Memento: Wayback Machine]
- Byrne, H., & Rarugal, C. (2019, June 6). Workshop: Reflecting on how we train new starters in web archiving. International Internet Preservation Coalition General Assembly and Web Archiving Conference, Zagreb, Croatia, 6–7 June 2019. Retrieved 2022-02-02, from <https://digital.library.unt.edu/ark:/67531/metadc1609017/>. [URL Memento: archive today]
- Byrne, H., & Rarugal, C. (2020, May 24). Reflecting on how we train new starters in web archiving [Blog post]. IIPC Blog. Retrieved 2020-05-24, from <https://netpreserveblog.wordpress.com/2020/05/24/reflecting-on-how-we-train-new-starters-in-web-archiving/>. [URL Memento: Wayback Machine]
- CAIN. (n.d.). PRONI Records on CAIN: About PRONI [Web page]. CAIN (Ulster University). <https://cain.ulster.ac.uk/proni/aboutproni.html> [URL Memento: Wayback Machine]
- Chakrabarti, S., Dom, B., Kumar, S. R., Raghavan, P., Rajagopalan, S., & Tomkins, A. (1999). Hypersearching the Web. *Scientific American*, 280(6), 54–60. <http://www.jstor.org/stable/26058288>. JSTOR
- Cho, J., & Garcia-Molina, H. (2000). The Evolution of the Web and Implications for an Incremental Crawler. *Proceedings of 26th International Conference on Very Large Data Bases, September 10-14, Cairo, Egypt, 2000*. Retrieved 2021-11-23, from <http://www.vldb.org/conf/2000/P200.pdf>. [URL Memento: Wayback Machine]
- Chudoba, B. (n.d.). How much time are respondents willing to spend on your survey? [Web page]. SurveyMonkey. Retrieved 2018-05-17, from [https://www.surveymonkey.com/curiosity/survey\\_completion\\_times/](https://www.surveymonkey.com/curiosity/survey_completion_times/). [URL Memento: Wayback Machine]
- Cocciolo, A. (2015). The Rise and Fall of Text on the Web: A Quantitative Study of Web Archives. *Information Research: An International Electronic Journal*, 20(3). Retrieved 2022-01-19, from <https://eric.ed.gov/?id=EJ1077827>. [URL Memento: Wayback Machine]
- CollEx - Persée. (n.d.). ResPaDon – CollEx - Persée [Web page]. CollEx - Persée. Retrieved 2021-10-15, from <https://www.collexpersee.eu/projet/respadon/>. [URL Memento: Wayback Machine]



- Collins, S. (2018). The National Library of Ireland. *Alexandria*, 28(3), 177–181.  
<https://doi.org/10.1177/0955749019878523>. DOI: 10.1177/0955749019878523
- Committee on the Records of Government. (1985). Committee on the Records of Government. Report Undertaken by Council on Library Resources, Inc., Washington, DC.; Social Science Research Council, Washington, DC.; American Council of Learned Societies, New York. Retrieved 2019-01-30, from <https://eric.ed.gov/?id=ED269018>. [URL Memento: Wayback Machine] ERIC Number: ED269018
- Computer Hope. (n.d.). TCP/IP [Web page]. Computer Hope. Retrieved 2021-12-09, from <https://www.computerhope.com/jargon/t/tcpip.htm>. [URL Memento: Wayback Machine]
- Consortium of National and University Libraries. (n.d.). Consortium of National and University Libraries (CONUL) [Website]. CONUL. Retrieved 2022-07-07, from <https://conul.ie/>. [URL Memento: Wayback Machine]
- Consortium of National and University Libraries. (2012). Submission by CONUL, In response to Copyright and Innovation: A Consultation Paper prepared by the Copyright Review Committee, For the Department of Jobs, Enterprise and Innovation. Ireland: Consortium of National and University Libraries. Retrieved 2022-05-26, from <https://enterprise.gov.ie/en/consultations/consultations-files/conul.pdf>. [URL Memento: Wayback Machine]
- CoolTool. (2017). 6-10 Minutes Is the Ideal Survey Length [Blog post]. CoolTool Blog. Retrieved from <https://web.archive.org/web/20201111232524/https://cooltool.com/blog/6-10-minutes-is-the-ideal-survey-length> [Web archive: Wayback Machine; source URL: <https://cooltool.com/blog/6-10-minutes-is-the-ideal-survey-length>; Timestamp: 2020-11-11 23:25:24]
- Copyright Review Committee. (2012). Copyright and Innovation: A Consultation Paper. Dublin, Ireland: Copyright Review Committee. Retrieved 2016-12-17, from <http://www.cearta.ie/wp-content/uploads/2012/02/CRC-Consultation-Paper.pdf>. [URL Memento: Wayback Machine]
- Copyright Review Committee. (2013). Modernising Copyright [A Report prepared by the Copyright Review Committee for the Department of Jobs, Enterprise and Innovation]. Copyright Review Committee. Retrieved 2016-05-28, from <http://www.cearta.ie/wp-content/uploads/2013/10/CRC-Report.pdf>. [URL Memento: Wayback Machine]
- Coram, R. G. (2015, July 24). Geo-location in the 2014 UK Domain Crawl. UK Web Archive Blog. Retrieved 2021-06-19, from <https://blogs.bl.uk/webarchive/2015/07/geo-location-in-the-2014-uk-domain-crawl.html> [URL Memento: Wayback Machine]
- Costa, M. (2021). Full-Text and URL Search Over Web Archives. In D. Gomes, E. Demidova, J. Winters, & T. Risse (Eds.), *The Past Web: Exploring Web Archives* (pp. 71–84). Cham, Switzerland: Springer. DOI: 10.1007/978-3-030-63291-5\_7
- Costa, M., & Silva, M. J. (2010). Understanding the information needs of web archive users. In J. Masanès, A. Rauber, & M. Spaniol (Eds.), *Proceedings of the 10th International Web Archiving Workshop (IWAW '10), Vienna, Austria, September 22-23, 2010*, 9–16. IWAW. Retrieved from

- <https://web.archive.org/web/20110723173820/http://www.iwaw.net/10/IWAW2010.pdf> [Web archive: Wayback Machine; source URL: <http://www.iwaw.net/10/IWAW2010.pdf>; Timestamp: 2011-07-23 17:38:20]
- Costea, M.-D. (2018). *Report on the Scholarly Use of Web Archives* [Report]. Aarhus, Denmark: NetLab. Retrieved 2019-08-30, from [http://netlab.dk/wp-content/uploads/2018/02/Costea\\_Report\\_on\\_the\\_Scholarly\\_Use\\_of\\_Web\\_Archives.pdf](http://netlab.dk/wp-content/uploads/2018/02/Costea_Report_on_the_Scholarly_Use_of_Web_Archives.pdf) [URL Memento: Wayback Machine]
- Cowls, J. (2013, November 15). Fahrenheit 401: Digital Deletion Is Incompatible with Democracy. Josh Cowls. Retrieved 2020-12-27, from <https://joshcowls.com/2013/11/15/fahrenheit-401-digital-deletion-is-incompatible-with-democracy/>. [URL Memento: Wayback Machine]
- Craigle, V., Retteen, A., & Keele, B. J. (2022). Ending Law Review Link Rot: A Plea for Adopting DOI. *Legal Reference Services Quarterly*, 0(0), 1–5. DOI: 10.1080/0270319X.2022.2089810
- Crocetti, P. (2019, October 31). Archive vs. Backup and why you need to know the differences [Web page]. TechTarget. Retrieved 2022-01-29, from <https://www.techtarget.com/searchdatabackup/tip/What-is-the-difference-between-archives-and-backups>. [URL Memento: Wayback Machine]
- Crowe, C. (2012, June 30). Ruin of Public Record Office marked loss of great archive. *The Irish Times*. Retrieved from <https://www.irishtimes.com/opinion/ruin-of-public-record-office-marked-loss-of-great-archive-1.1069843> [URL Memento: Wayback Machine]
- Darcy, E., Charthaigh, C. N., Lalor, M., Sinnott, J., & Shank, C. (Eds.). (2021). In Her Shoes: Stories of the Eighth Amendment (Ireland). Digital Repository of Ireland, Collections. DOI: 10.7486/DRI.wm11nd02p
- Dawson, C. (2019). *A-Z of Digital Research Methods* (Paperback). Oxon; New York: Routledge.
- Day, M. (2003). *Collecting and preserving the World Wide Web: A feasibility study undertaken for the JISC and Wellcome Trust* (Version 1.0). UKOLN, University of Bath. [https://web.archive.org/web/20030408060243/http://www.jisc.ac.uk/uploaded\\_documents/archiving\\_feasibility.pdf](https://web.archive.org/web/20030408060243/http://www.jisc.ac.uk/uploaded_documents/archiving_feasibility.pdf). [Web archive: Wayback Machine; source URL: [http://www.jisc.ac.uk/uploaded\\_documents/archiving\\_feasibility.pdf](http://www.jisc.ac.uk/uploaded_documents/archiving_feasibility.pdf); Timestamp: 2003-04-08 06:02:43]
- Day, M. (2006). The Long-Term Preservation of Web Content. In J. Masanès (Ed.), *Web Archiving* (pp. 177–199). Berlin, Heidelberg: Springer-Verlag.
- De Haan, T. (2018, November 29). Bit by bit, byte by byte: Web archaeology going strong in the Netherlands! [Web page]. DPC Blog. Retrieved 2021-10-11, from <https://www.dpconline.org/blog/wdpd/bit-by-bit-byte-by-byte>. [URL Memento: Wayback Machine]
- De Haan, T., Jansma, R., & Vogel, P. (2017). *DIY Handboek voor Webarcheologie* [Guide]. Amsterdam Museum. Retrieved 2021-09-16, [https://hart.amsterdam/image/2017/11/17/20171116\\_freeze\\_diy\\_handboek.pdf](https://hart.amsterdam/image/2017/11/17/20171116_freeze_diy_handboek.pdf). [URL Memento: Wayback Machine]

- Dellavalle, R. P., Hester, E. J., Heilig, L. F., Drake, A. L., Kuntzman, J. W., Graber, M., & Schilling, L. M. (2003). Going, Going, Gone: Lost Internet References. *Science*, 302(5646), 787–788. American Association for the Advancement of Science. DOI: 10.1126/science.1088234
- Denev, D., Mazeika, A., Spaniol, M., & Weikum, G. (2009). SHARC: Framework for Quality-conscious Web Archiving. *Proceedings of the VLDB Endowment*, 2, 586–597. DOI: 10.14778/1687627.1687694
- Denev, D., Mazeika, A., Spaniol, M., & Weikum, G. (2011). The SHARC Framework for Data Quality in Web Archiving. *The VLDB Journal*, 20(2), 183–207. DOI: 10.1007/s00778-011-0219-9
- Denning, S. (2011, July 23). How Do You Change An Organizational Culture? Forbes. Retrieved 2014-11-16, from <https://www.forbes.com/sites/stevedenning/2011/07/23/how-do-you-change-an-organizational-culture/>. [URL Memento: Wayback Machine]
- Department of Arts, Heritage, Regional, Rural and Gaeltacht Affairs. (2017a, April 21). Consultation on the Legal Deposit of published digital material in the 21st century in the context of Copyright legislation [Archived web page]. Department of Arts, Heritage, Regional, Rural and Gaeltacht Affairs, 21 April 2017. Retrieved from <https://web.archive.org/web/20170424201229/http://www.ahrrga.gov.ie/consultation-on-the-legal-deposit-of-published-digital-material-in-the-21st-century-in-the-context-of-copyright-legislation/>. [Web archive: Wayback Machine; source URL: <http://www.ahrrga.gov.ie/consultation-on-the-legal-deposit-of-published-digital-material-in-the-21st-century-in-the-context-of-copyright-legislation/>; Timestamp: 2017-04-24 20:12:29]
- Department of Arts, Heritage, Regional, Rural and Gaeltacht Affairs. (2017b). Consultation on the Legal Deposit of published digital material in the 21st century in the context of Copyright legislation [Archived web page]. Department of Arts, Heritage, Regional, Rural and Gaeltacht Affairs, 21 April 2017. Retrieved from <https://web.archive.org/web/20170424204921/http://www.ahrrga.gov.ie/app/uploads/2017/04/consultation-on-the-legal-deposit-of-published-digital-material-in-the-21st-century-in-the-context-of-copyright-legislation.pdf>. [Web archive: Wayback Machine; source URL: <http://www.ahrrga.gov.ie/app/uploads/2017/04/consultation-on-the-legal-deposit-of-published-digital-material-in-the-21st-century-in-the-context-of-copyright-legislation.pdf>; Timestamp: 2017-04-24 20:49:21]
- Department of Culture, Heritage and the Gaeltacht. (2017, December). Department of Culture, Heritage, and the Gaeltacht response to submissions received in respect of the consultation on the Legal Deposit of published digital material in the 21st century in the context of Copyright Legislation. Department of Culture, Heritage and the Gaeltacht, December 2017. Retrieved from <https://web.archive.org/web/20180305230900/https://www.chg.gov.ie/app/uploads/2017/12/public-consultation-reponse-summary-document-v2.pdf>. [Web archive: Wayback Machine; source URL: <https://www.chg.gov.ie/app/uploads/2017/12/public-consultation-reponse-summary-document-v2.pdf>; Timestamp: 2018-03-05 23:09:00]
- Department of Enterprise, Trade and Employment. (n.d.). Dara Calleary [Web page]. Department of Enterprise, Trade and Employment. Retrieved 2022-10-31, from

- <https://enterprise.gov.ie/en/who-we-are/ministers/dara-calleary.html>. [URL Memento: Wayback Machine]
- Department of Enterprise, Trade and Employment. (n.d.). Robert Troy [Archived web page]. Department of Enterprise, Trade and Employment. Retrieved from <https://web.archive.org/web/20220823133750/https://enterprise.gov.ie/en/Who-We-Are/Ministers/Robert-Troy/> [Web archive: Wayback Machine; source URL: <https://enterprise.gov.ie/en/Who-We-Are/Ministers/Robert-Troy/>; Timestamp: 2022-08-23 13:37:50]
- Department of Enterprise, Trade and Employment. (2016, August 4). General Scheme of a Copyright Bill approved by Government [Web page]. Department of Enterprise, Trade and Employment. Retrieved 2021-05-06, from <https://enterprise.gov.ie/en/News-And-Events/Department-News/2016/August/04082016a.html>. [URL Memento: Wayback Machine]
- Department of Enterprise, Trade and Employment. (2021, December 5). Ministers—DETE [Archived web page]. Department of Enterprise, Trade and Employment. Retrieved from <https://web.archive.org/web/20211205081704/https://enterprise.gov.ie/en/Who-We-Are/Ministers/>. [Web archive: Wayback Machine; source URL: <https://enterprise.gov.ie/en/Who-We-Are/Ministers/>; Timestamp: 2021-12-05 08:17:04]
- Department of Jobs, Enterprise and Innovation. (2012, June 13). Submissions Received 2012 on foot of the Copyright Review Consultation Paper [Archived web page]. Department of Jobs, Enterprise and Innovation. Retrieved from [https://wayback.archive-it.org/org-1444/20120613230622/http://www.djei.ie/science/ipr/crc\\_submissions2.htm](https://wayback.archive-it.org/org-1444/20120613230622/http://www.djei.ie/science/ipr/crc_submissions2.htm). [Web archive: NLI Web Archive; source: URL: Timestamp: 2012-06-13 23:06:22]
- Department of Jobs, Enterprise, and Innovation. (2012, May 31). Consultation on the Review of the Copyright and Related Rights Act 2000 [Archived web page]. Department of Jobs, Enterprise, and Innovation. Retrieved from [https://wayback.archive-it.org/org-1444/20120613230629/http://www.djei.ie/science/ipr/copyright\\_review\\_2011.htm](https://wayback.archive-it.org/org-1444/20120613230629/http://www.djei.ie/science/ipr/copyright_review_2011.htm) [Web archive: NLI Web Archive; source: URL: [http://www.djei.ie/science/ipr/crc\\_submissions2.htm](http://www.djei.ie/science/ipr/crc_submissions2.htm); Timestamp: 2012-06-13 23:06:22]
- Department of Jobs, Enterprise, and Innovation. (2012, May 29). Copyright Review Committee announces limited extension of time for submissions [Archived web page]. Department of Jobs, Enterprise, and Innovation. Retrieved from <https://web.archive.org/web/20120603221249/http://www.djei.ie/press/2012/20120529a.htm> . [Web archive: Wayback Machine; source URL: <http://www.djei.ie/press/2012/20120529a.htm>; Timestamp: 2012-06-03 22:12:49]
- Department of Jobs, Enterprise, and Innovation. (2012, March 2). Submissions Received by the Copyright Review Committee [Archived web page]. Department of Jobs, Enterprise, and Innovation. Retrieved from [https://web.archive.org/web/20120618012622/http://www.djei.ie/science/ipr/crc\\_submissions.htm](https://web.archive.org/web/20120618012622/http://www.djei.ie/science/ipr/crc_submissions.htm). [Web archive: NLI Web Archive; source: URL:

- [http://www.djei.ie/science/ipr/crc\\_submissions2.htm](http://www.djei.ie/science/ipr/crc_submissions2.htm); Timestamp: 2012-06-13 23:06:22]
- Digital Repository Ireland. (2022, September 16). DRI Project Archiving Reproductive Health Wins Digital Preservation Award. Digital Repository Ireland. <https://www.dri.ie/dri-project-archiving-reproductive-health-wins-digital-preservation-award> . [URL Memento: Wayback Machine]
- Dollar, C. (1978). Appraising Machine-Readable Records. *The American Archivist*, 41(4), 423–430. <https://doi.org/10.17723/aarc.41.4.g333h26662621363>. DOI: 10.17723/aarc.41.4.g333h26662621363
- Dougherty, M. (2007). Archiving the Web: Collection, Documentation, Display, and Shifting Knowledge Production Paradigms [PhD Dissertation, University of Washington]. ProQuest One Academic. ProQuest document ID: 304794229
- Dougherty, M., & Heuvel, C. van den. (2009, April). Historical Infrastructures for Web Archiving: Annotation of Ephemeral Collections for Researchers and Cultural Heritage Institutions. *Massachusetts Institute of Technology 6th Media in Transition Conference, Cambridge, Massachusetts, 2009*. Massachusetts, USA. Retrieved from [https://web.archive.org/web/20120616174354/http://web.mit.edu/comm-forum/mit6/papers/Dougherty\\_Heuvel.pdf](https://web.archive.org/web/20120616174354/http://web.mit.edu/comm-forum/mit6/papers/Dougherty_Heuvel.pdf) [Web Archive: Wayback Machine; source URL: [http://web.mit.edu/comm-forum/mit6/papers/Dougherty\\_Heuvel.pdf](http://web.mit.edu/comm-forum/mit6/papers/Dougherty_Heuvel.pdf); Timestamp: 2012-06-16 17:43:54]
- Dougherty, M., Meyer, E. T., Madsen, C., McCarthy, van den Heuvel, C., Thomas, A., & Wyatt, S. (2010). *Researcher Engagement with Web Archives: State of the Art* (Joint Information Systems Committee Report, August 2010). London: Joint Information Systems Committee (JISC). Retrieved 2020-07-31, from [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1714997](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1714997). [URL Memento: Wayback Machine]
- Drucker, J. (2012). Humanistic Theory and Digital Scholarship. In Matthew K. Gold (Ed.), *Debates in the Digital Humanities*. Minneapolis: University of Minnesota Press. DOI: 10.5749/minnesota/9780816677948.003.0011
- Dupont, H. (1999). Legal Deposit in Denmark—The New Law and Electronic Products. *LIBER Quarterly: The Journal of the Association of European Research Libraries*, 9(2), 244–251. DOI: 10.18352/lq.7539
- Eldakar, Y., & Holownia, O. (2022, June 29). Republishing IIPC Collections Through Alternative Interfaces. *IIPC General Assembly and Web Archiving Conference, Library of Congress, May 2022* (IIPC WAC 2022). Retrieved from <https://www.youtube.com/watch?v=15AGoOJBM6E> [YouTube]
- Eltgroth, D. (2009). Best Evidence and the Wayback Machine: Toward a Workable Authentication Standard for Archived Internet Evidence. *Fordham Law Review*, 78(1), 181. Retrieved 2021-04-28, from <https://ir.lawnet.fordham.edu/flr/vol78/iss1/5>. [URL Memento: Wayback Machine]
- Engholm, I. (2002). Digital style history: The development of graphic design on the Internet. *Digital Creativity*, 13(4), 193–211. DOI: 10.1076/digc.13.4.193.8672

- European Commission. (n.d.). Data protection [Web page]. European Commission. Retrieved 2021-11-30, from [https://ec.europa.eu/info/law/law-topic/data-protection\\_en](https://ec.europa.eu/info/law/law-topic/data-protection_en). [URL Memento: Wayback Machine]
- Farrell, M., McCain, E., Praetzellis, M., Thomas, G., & Walker, P. (2018). *Web Archiving in the United States: A 2017 Survey* [NDSA Report]. USA: National Digital Stewardship Alliance (NDSA). Retrieved 2021-02-25, from <https://osf.io/r5pqk/>. DOI 10.17605/OSF.IO/R5PQK [URL Memento: Archive.today]
- Feather, J. (1994). *Publishing, piracy, and politics: An historical study of copyright in Britain*. New York, N.Y: Mansell.
- Fetterly, D., Manasse, M., Najork, M., & Wiener, J. (2003). A large-scale study of the evolution of web pages. *Proceedings of the 12th International Conference on World Wide Web (WWW '03)*, 669–678. DOI: 10.1145/775152.775246
- Fishbein, M. H. (1972). Appraising Information in Machine Language Form. *The American Archivist*, 35(1), 35–43. JSTOR ID: 40291594
- Foot, K. A., & Schneider, S. M. (2006). *Web Campaigning*. Cambridge, MA, USA: MIT Press.
- Gardner, J. B. (1997). Report on Documenting the Digital Age Conference. History Associates Incorporated. Retrieved from <https://web.archive.org/web/20180524134523/https://www.historyassociates.com/wp-content/uploads/2015/09/Documenting-the-Digital-Age.pdf>. [Web archive: Wayback Machine; source URL: Timestamp: 2018-05-24 13:45:23]
- Gataveckaitė, G. (2022, March 14). Sinn Féin deletes thousands of statements from its website due to 'outdated content'. *The Irish Independent*. <https://www.independent.ie/irish-news/politics/sinn-fein-deletes-thousands-of-statements-from-its-website-due-to-outdated-content-41443385.html>. [URL Memento: Wayback Machine]
- Germain, C. A. (2000). URLs: Uniform resource locators or unreliable reliable resource locators? *College and Research Libraries*, 61(4), 359–365. College & Research Libraries. DOI: 10.5860/crl.61.4.359 [URL Memento: Archive.today]
- Gillies, J., & Cailliau, R. (2000). *How the Web was Born: The Story of the World Wide Web*. Oxford, New York: Oxford University Press.
- Goh, D. H.-L., & Ng, P. K. (2007). Link decay in leading information science journals. *Journal of the American Society for Information Science and Technology*, 58(1), 15–24. DOI: 10.1002/asi.20513
- Gomes, D., & Costa, M. (2014). The Importance of Web Archives for Humanities. *International Journal of Humanities & Arts Computing: A Journal of Digital Humanities*, 8(1), 106–123. DOI: 10.3366/ijhac.2014.0122
- Gomes, D., Demidova, E., Winters, J., & Risse, T. (2021a). Preface. In D. Gomes, E. Demidova, J. Winters, & T. Risse (Eds.), *The Past Web: Exploring Web Archives* (pp. xi–xiii). Cham, Switzerland: Springer.
- Gomes, D., Demidova, E., Winters, J., & Risse, T. (Eds.). (2021b). *The Past Web: Exploring Web Archives*. Cham, Switzerland: Springer.
- Gomes, D., Miranda, J., & Costa, M. (2011). A Survey on Web Archiving Initiatives. In S. Gradmann, F. Borri, C. Meghini, & H. Schuldt (Eds.), *Research and Advanced Technology for Digital Libraries*. (Vol. 6966). DOI: 10.1007/978-3-642-24469-8\_41

- Gooding, P., Terras, M., & Berube, L. (2019). Towards User-Centric Evaluation of UK Non-Print Legal Deposit: A Digital Library Futures White Paper. *Digital Library Futures*.  
<https://digitalcommons.unl.edu/scholcom/180> [URL Memento: Wayback Machine]
- Gooding, P., Terras, M., & Berube, L. (2021). Identifying the future direction of legal deposit in the United Kingdom: The Digital Library Futures approach. *Journal of Documentation*, 77(5), 1154–1172. DOI: 10.1108/JD-09-2020-0159
- Gorsky, M. (2015). Into the Dark Domain: The UK Web Archive as a Source for the Contemporary History of Public Health. *Social History of Medicine*, 28(3), 596–616. DOI: 10.1093/shm/hkv028
- Graham, P. M. (2017). Guest Editorial: Reflections on the Ethics of Web Archiving. *Journal of Archival Organization*, 14(3–4), 103–110. DOI: 10.1080/15332748.2018.1517589
- Graham, P. S. (1994). Intellectual Preservation: Electronic Preservation of the Third Kind. Commission on Preservation and Access. Retrieved 2021-05-07, from  
<https://www.clir.org/pubs/reports/graham/intpres/>. [URL Memento: Wayback Machine] ERIC Number: ED369414
- Greene, D. (2020, September 21). Exploring Web Archive Networks: The Case of the 2018 Irish Presidential Election [Conference presentation (video; abstract)]. *Engaging with Web Archives: 'Opportunities, Challenges and Potentialities', (#EWAVirtual), Maynooth University Arts and Humanities Institute, Co. Kildare, Ireland, [online], 21-22 September 2022*. DOI: 10.5281/zenodo.4058013 [URL Memento: Wayback Machine] YouTube:  
[https://www.youtube.com/watch?v=ld\\_fXefv-5E&t=1s](https://www.youtube.com/watch?v=ld_fXefv-5E&t=1s)
- Greene, D., & Ryan, M. (2019, June 19). Exploring Selective Web Archives via Network Analysis: An Irish Case Study [Poster]. *LIBER Annual Conference, Dublin, Ireland, June 2019*. DOI: 10.5281/zenodo.3250157. [URL Memento: Wayback Machine]
- Grotke, A. (2011, December). FEATURE: Web Archiving at the Library of Congress. *Information Today*. Retrieved 2018-07-21, from  
<https://www.infotoday.com/cilmag/dec11/Grotke.shtml>. [URL Memento: Wayback Machine]
- Harter, S. P., & Kim, H. J. (1996). Electronic journals and scholarly communication: A citation and reference study. *Information Research*, 2(1). Retrieved 2021-02-26, from  
<http://informationr.net/ir/2-1/paper9a.html>. [URL Memento: Wayback Machine]
- Healy, S. (2016, October 27). Here today, gone tomorrow: A case study on the necessity for a more rigorous approach to the preservation of online Irish cultural and political heritage. *Institutions and Ireland: Public Cultures*, Trinity College Dublin, Ireland, 27 October 2016. Retrieved 2022-04-25, from <https://hcommons.org/deposits/item/hc:45365/>. DOI: 10.17613/xy5v-6b63. [URL Memento: Wayback Machine]
- Healy, S. (2021, April). Fleeting narratives: Web histories of Irish LGBT activism [Conference abstract]. *Women's History Association of Ireland Annual Conference 2020/2021: Besieged bodies: Gendered violence, sexualities and motherhood*, University College Dublin, March 2021. Retrieved 2021-02-26, from <https://womenshistoryassociation.com/whai-conference-2020-2021/>. [URL Memento: Wayback Machine]

- Healy, S., Byrne, H., Schmid, K., Boté-Vericad, J.-J., Floody, L. (2023). Towards a Glossary for Web Archive Research: Version 1.0. *WARCnet Special Report*. Aarhus, Denmark: WARCnet.
- Healy, S., Byrne, H., Schmid, K., Floody, L., Boté-Vericad, J.-J. (2021+) Zotero - Groups - Towards a Glossary for Web Archive Research.  
Zotero, [https://www.zotero.org/groups/4380600/towards\\_a\\_glossary\\_for\\_web\\_archive\\_research](https://www.zotero.org/groups/4380600/towards_a_glossary_for_web_archive_research). [export files available in OSF: <https://osf.io/vf7gt/>]
- Hendler, J. (2003). Science and the Semantic Web. *Science*, 299(5606), 520–521. DOI: 10.1126/science.1078874
- Hester, E. J., Heilig, L. F., Drake, A. L., Johnson, K. R., Vu, C. T., Schilling, L. M., & Dellavalle, R. P. (2004). Internet Citations in Oncology Journals: A Vanishing Resource? *Journal of the National Cancer Institute*, 96(12), 969–971. DOI: 10.1093/jnci/djh181
- Hindley, M. (2013, August). The Rise of the Machines. *HUMANITIES: The Magazine of The National Endowment for the Humanities*, 34(4). Retrieved 2017-01-26, from <https://www.neh.gov/humanities/2013/julyaugust/feature/the-rise-the-machines>
- Hockey, S. (2004). The History of Humanities Computing. In S. Schreibman, R. Siemens, & J. Unsworth (Eds.), *A Companion to Digital Humanities* (Web version). Oxford: Blackwell Publishing Professional. Retrieved 2018-06-28, from [http://digitalhumanities.org:3030/companion/view?docId=blackwell/9781405103213/9781405103213.xml&chunk.id=ss1-2-1&toc.depth=1&toc.id=ss1-2-1&brand=9781405103213\\_brand](http://digitalhumanities.org:3030/companion/view?docId=blackwell/9781405103213/9781405103213.xml&chunk.id=ss1-2-1&toc.depth=1&toc.id=ss1-2-1&brand=9781405103213_brand). [URL Memento: Wayback Machine]
- Hockx-Yu, H. (2014). Access and Scholarly Use of Web Archives. *Alexandria: The Journal of National and International Library and Information Issues*, 25(11), 113–127. DOI: 10.7227/ALX.0023
- Hofheinz, A. (2010). A History of Allah.com. In N. Brügger (Ed.), *Web History* (pp. 105–135). New York: Peter Lang.
- Holzmann, H., & Nejdli, W. (2021). A Holistic View on Web Archives. In D. Gomes, E. Demidova, J. Winters, & T. Risse (Eds.), *The Past Web: Exploring Web Archives* (pp. 85–99). Cham, Switzerland: Springer.
- Holzmann, H., Nejdli, W., & Anand, A. (2016). The Dawn of Today's Popular Domains: A Study of the Archived German Web over 18 Years. *Proceedings of the 16th ACM/IEEE-CS on Joint Conference on Digital Libraries*, pp. 73–82. DOI: 10.1145/2910896.2910901
- Huc-Hepher, S., & Wells, N. (2021). Exploring Online Diasporas: London's French and Latin American Communities in the UK Web Archive. In D. Gomes, E. Demidova, J. Winters, & T. Risse (Eds.), *The Past Web: Exploring Web Archives* (pp. 189–201). Cham, Switzerland: Springer. DOI: 10.1007/978-3-030-63291-5\_15
- Huurdemann, H. C., & Kamps, J. (2017). A Collaborative Approach to Research Data Management in a Web Archive Context. In Filip Kruse & Jesper Boserup Thestrup (Eds.), *Research Data Management - A European Perspective* (pp. 55–78). Berlin/Boston: De Gruyter Saur. DOI: 10.1515/9783110365634-005
- .IE Domain Registry. (2021). .IE Domain Profile Report 2021. .IE Domain Registry. Retrieved 2022-06-30, from <https://www.weare.ie/wp-content/uploads/2022/01/IE-DPR-2021.pdf>. [URL Memento: Wayback Machine]



International Internet Preservation Consortium. (n.d.). About the IIPC [Web page]. International Internet Preservation Consortium. Retrieved 2021-04-23, from <https://netpreserve.org/about-us/>. [URL Memento: Wayback Machine]

International Internet Preservation Consortium. (n.d.). Bibliography [Web page]. International Internet Preservation Consortium. Retrieved 2021-09-09, from <https://netpreserve.org/web-archiving/bibliography/>. [URL Memento: Wayback Machine]

International Internet Preservation Consortium. (n.d.). Collection development policies [Web page]. International Internet Preservation Consortium. Retrieved 2021-09-08, from <https://netpreserve.org/web-archiving/collection-development-policies/>. [URL Memento: Wayback Machine]

International Internet Preservation Consortium. (n.d.). IIPC Members [Web page]. International Internet Preservation Consortium. Retrieved 2021-11-23, from <https://netpreserve.org/about-us/members/>. [URL Memento: Wayback Machine]

International Internet Preservation Consortium. (n.d.). IIPC WAC 2022 program [Web page]. International Internet Preservation Consortium. Retrieved 2022-05-20, from <https://netpreserve.org/ga2022/wac/>. [URL Memento: Wayback Machine]

International Internet Preservation Consortium. (n.d.). International Internet Preservation Consortium – Home [Website]. International Internet Preservation Consortium. Retrieved 2021-04-23, from <https://netpreserve.org/>. [URL Memento: Wayback Machine]

International Internet Preservation Consortium. (n.d.). International Internet Preservation Consortium - Archive-It [Web page]. Archive-It. Retrieved 2022-06-03, from <https://archive-it.org/home/IIPC>. [URL Memento: Wayback Machine]

International Internet Preservation Consortium. (n.d.). International Internet Preservation Consortium—Welcome (2004). International Internet Preservation Consortium. Retrieved from <https://web.archive.org/web/20040603014115/http://netpreserve.org/about/index.php>. [Web archive: Wayback Machine; source URL: <http://netpreserve.org/about/index.php>; Timestamp: 2004-06-03 01:41:15]

International Internet Preservation Consortium. (n.d.). Legal deposit [Web page]. International Internet Preservation Consortium. Retrieved 2021-11-03, from <https://netpreserve.org/web-archiving/legal-deposit/>. [URL Memento: Wayback Machine]

International Internet Preservation Consortium. (n.d.b). The WARC Format 1.0 [Web page]. IIPC/Github.io. Retrieved 2021-07-06, from <https://iipc.github.io/warc-specifications/specifications/warc-format/warc-1.0/>. [URL Memento: Wayback Machine]

International Internet Preservation Consortium - curators email list. (2017, April 4). Call for Participation: Archives Unleashed 4.0: Web Archive Datathon [distribution list communication].

Internet Archive Help Center. (n.d.). Wayback Machine General Information [Web page]. Internet Archive Help Center. Retrieved 2022-08-21, from

- <https://help.archive.org/help/wayback-machine-general-information/>. [URL Memento: Wayback Machine]
- Irish Manuscripts Commission. (n.d.) Mission. Irish Manuscripts Commission. Retrieved 2022-09-26, from <https://www.irishmanuscripts.ie/about-us/mission/>. [URL Memento: Wayback Machine]
- Jackson, A., Lin, J., Milligan, I., & Ruest, N. (2016). Desiderata for Exploratory Search Interfaces to Web Archives in Support of Scholarly Activities. *Proceedings of the 16th ACM/IEEE-CS on Joint Conference on Digital Libraries*, 103–106. DOI: 10.1145/2910896.2910912
- Jackson, A. (2015a, April 27). Ten years of the UK web archive: What have we saved? [Conference presentation (slides)]. International Internet Preservation Consortium, General Assembly, Palo Alto, 27 April 2015, Palo Alto. Retrieved 2020-12-05, from <https://blogs.bl.uk/webarchive/2015/09/ten-years-of-the-uk-web-archive-what-have-we-saved.html>. [URL Memento: Wayback Machine]
- Jackson, A. (2015b, November 20). The Provenance of Web Archives [Blog post]. UK Web Archive Blog. Retrieved 2022-02-24, from <https://britishlibrary.typepad.co.uk/webarchive/2015/11/>. [URL Memento: Wayback Machine]
- Jackson, A. (2022, January 6). UKWA 2021 Technical update [Blog post]. UK Web Archive Blog. Retrieved 2022-01-07, from <https://blogs.bl.uk/webarchive/2022/01/ukwa-2021-technical-update.html>. [URL Memento: Wayback Machine]
- Jatowt, A., Kawai, Y., Ohshima, H., & Tanaka, K. (2008). What can history tell us? Towards different models of interaction with document histories. *Proceedings of the Nineteenth ACM Conference on Hypertext and Hypermedia*, 5–14. DOI: 10.1145/1379092.1379098
- JSTOR. (n.d.). *About JSTOR* [Web page]. JSTOR. Retrieved 2019-09-05, <https://about.jstor.org/>. [URL Memento: Wayback Machine]
- Kahn, R. (2019). The Nation is in the network. In N. Brügger & D. Laursen (Eds.), *The Historical Web and Digital Humanities: The Case of National Web Domains* (pp. 161–177). London & New York: Routledge.
- Kelle, U. (1995). Introduction: An Overview of Computer-aided Methods in Qualitative Research. In U. Kelle, Geraldine Prein, & K. Bird (Eds.), *Computer-Aided Qualitative Data Analysis: Theory, Methods and Practice* (pp. 1–18). London: SAGE Publications.
- Klein, M., Sompel, H. V. de, Sanderson, R., Shankar, H., Balakireva, L., Zhou, K., & Tobin, R. (2014). Scholarly Context Not Found: One in Five Articles Suffers from Reference Rot. *PLoS ONE*, 9(12), e115253. DOI: 10.1371/journal.pone.0115253
- Koehler, W. (1999). Digital Libraries and World Wide Web Sites and Page Persistence. *Information Research*, 4(4). Information Research. Retrieved 2021-02-11, from <http://www.informationr.net/ir/4-4/paper60.html>. [URL Memento: Wayback Machine]
- Koehler, W. (2015). *Ethics and Values in Librarianship: A History*. New York & London: Rowman & Littlefield.
- Koerbin, P. (2013a, July 2). Web Archiving in a Fast Moving World: The Tricky Task of Capturing Websites While the News Is Hot [Blog post]. National Library of Australia Blog. Retrieved from <https://web.archive.org/web/20150910044329/https://www.nla.gov.au/australias-web->

- archives/2013/07/02/web-archiving-in-a-fast-moving-world. [Web archive: Wayback Machine; source: URL; Timestamp: 2015-09-10 04:43:29]
- Koerbin, P. (2013b, August 9). Archiving Online Election Campaigns: Explaining the Process Behind How We Preserve Australia's Election Websites [Blog post]. National Library of Australia Blog. Retrieved from <https://web.archive.org/web/20150905084451/https://www.nla.gov.au/australias-web-archives/2013/08/09/archiving-online-election-campaigns> [Web archive: Wayback Machine; source URL: <https://www.nla.gov.au/australias-web-archives/2013/08/09/archiving-online-election-campaigns>; Timestamp: 2015-09-05 08:44:51]
- Koerbin, P. (2021). National Web Archiving in Australia: Representing the Comprehensive. In D. Gomes, E. Demidova, J. Winters, & T. Risse (Eds.), *The Past Web: Exploring Web Archives* (pp. 23–32). Cham, Switzerland: Springer. DOI: 10.1007/978-3-030-63291-5\_3
- Kuny, T. (1997, September 4). A Digital Dark Ages? Challenges in the Preservation of Electronic Information. *63rd IFLA General Conference, Copenhagen, August 31 - September 5, 1997*. Retrieved 2021-07-23, from <http://archive.ifla.org/IV/ifla63/63kuny1.pdf>. [URL Memento: Wayback Machine]
- Lancaster, F. W. (1995). The Evolution of Electronic Publishing. *Library Trends*, 43(4). Retrieved 2022-07-26, from <https://www.ideals.illinois.edu/items/7940>. [URL Memento: Wayback Machine] Permalink: <https://hdl.handle.net/2142/7981>
- Larivière, J. (2000). *Guidelines for legal deposit legislation* (UNESCO) [Programme and meeting document]. United Nations Educational, Scientific and Cultural Organization. Retrieved 2019-04-25, from <https://unesdoc.unesco.org/ark:/48223/pf0000121413>. Document code: CII.00/WS/7
- Lawrence, S., & Giles, C. L. (1999). Accessibility of information on the web. *Nature*, 400(6740), 107–107. DOI: 10.1038/21987 [URL Memento: Wayback Machine]
- Lawrence, S., Pennock, D. M., Flake, G. W., Krovetz, R., Coetzee, F. M., Glover, E., Nielsen, F. Å., Kruger, K., & Giles, C. L. (2001). Persistence of Web References in Scientific Research. *Computer*, 34(1), 26–31. IEEE Xplore. DOI: 10.1109/2.901164
- Lay, P. (2017, May 5). History has a History. *History Today*, 67(5). Retrieved 2020-09-04, from <https://www.historytoday.com/archive/editor/history-has-history> [URL Memento: Wayback Machine]
- Leiner, B. M., Cerf, V. G., Clark, D. D., Kahn, R. E., Kleinrock, L., Lynch, D. C., Postel, J., Roberts, L. G., & Wolff, S. (1997). *Brief History of the Internet*. Internet Society. Retrieved 2018-03-29, from [https://cdn.prod.internetsociety.org/wp-content/uploads/2017/09/ISOC-History-of-the-Internet\\_1997.pdf](https://cdn.prod.internetsociety.org/wp-content/uploads/2017/09/ISOC-History-of-the-Internet_1997.pdf). [URL Memento: Wayback Machine]
- Leetaru, K. (2017, January 13). Why Aren't We Doing More with Our Web Archives? *Forbes*, AI & Big Data. Retrieved 2021-01-24, from <https://www.forbes.com/sites/kalevleetaru/2017/01/13/why-arent-we-doing-more-with-our-web-archives/?sh=155f433c498a>. [URL Memento: Archive.today]
- Leetaru, K. (2019, May 7). Why Web Archives Need To Engage With Researchers. *Forbes*, AI & Big Data. Retrieved 2019-11-30, from

- <https://www.forbes.com/sites/kalevleetaru/2019/05/07/why-web-archives-need-to-engage-with-researchers/>. [URL Memento: Wayback Machine]
- Library of Congress. (n.d.). ARC\_IA, Internet Archive ARC file format [Web page]. Sustainability of Digital Formats: Planning for Library of Congress Collections. Retrieved 2021-12-18, from <https://www.loc.gov/preservation/digital/formats/fdd/fdd000235.shtml>
- Library of Congress, Public Affairs Office. (1998, October 13). Alexa Internet Donates Archive of the World Wide Web to Library of Congress: First Large-Scale Digital Donation Ensures Preservation of Digital Cultural Artifacts [web version]. News from the Library of Congress, PR 98-167. Retrieved from <https://web.archive.org/web/20030423175610/http://www.loc.gov/today/pr/1998/98-167.html>. [Web archive: Wayback Machine; source URL: <http://www.loc.gov/today/pr/1998/98-167.html>; Timestamp: 2003-04-23 17:56:10]
- Library of Trinity College Dublin. (n.d.a). Electronic Legal Deposit: Accessing UK eLD content [Web page]. Trinity College Dublin. Retrieved 2022-09-28, from <https://libguides.tcd.ie/c.php?g=691904&p=4957591>. [URL Memento: Archive.Today]
- Library of Trinity College Dublin. (n.d.b). Electronic Legal Deposit: History and Background [Web page]. Trinity College Dublin. Retrieved 2022-09-28, from <https://libguides.tcd.ie/c.php?g=691904&p=4957589>. [URL Memento: Wayback Machine]
- Lombord, S. (2019). Ethical Considerations for Web Archives and Web History Research. In N. Brügger & I. Milligan (Eds.), *The SAGE Handbook of Web History* (pp. 99-111). London: SAGE Publications.
- Lowcock, J. (2020, February 26). Guide: How to Delete Your Site from the Internet Archive (Wayback Machine / Archive.org) [Digital-media; personal]. Joshua Lowcock, New York. Retrieved 2020-05-26, from <https://www.joshualowcock.com/tips-tricks/how-to-delete-your-site-from-the-internet-archive-wayback-machine-archive-org/>. [URL Memento: Wayback Machine]
- Lyman, P. (2002). Archiving the World Wide Web. In Council on Library and Information Resources & Library of Congress (Eds.), *Building a National Strategy for Preservation: Issues in Digital Media Archiving* (Online/pdf, pp. 38–51). USA: Council on Library and Information Resources and the Library of Congress. Retrieved 2021-03-29, from <https://www.clir.org/wp-content/uploads/sites/6/pub106.pdf>. [URL Memento: Wayback Machine]
- Mackinnon, K. (2020). DELETE MY ACCOUNT: Ethical Approaches to Researching Youth Cultures in Historical Web Archives. *Engaging with Web Archives: 'Opportunities, Challenges and Potentialities', (#EWAVirtual), Maynooth University Arts and Humanities Institute, Co. Kildare, Ireland, [online], 21-22 September 2022*. Retrieved from <https://www.youtube.com/watch?v=2lBX1Om6GLM>
- Mackinnon, K. (2021). Ethical Approaches to Youth Data in Historical Web Archives (Dispatch). *Studies in Social Justice*, 15(3), 442–449. DOI: 10.26522/ssj.v15i3.2541
- Maemura, E. (2018). What's cached is prologue: Reviewing recent web archives research towards supporting scholarly use. *Proceedings of the Association for Information Science and Technology*, 55(1), 327–336. DOI: 10.1002/pa2.2018.14505501036

- Maemura, E. (2022). Towards an Infrastructural Description of Archived Web Data. *WARCnet Papers*. Aarhus, Denmark: WARCnet. Retrieved 2022-05-09, from [https://cc.au.dk/fileadmin/dac/Projekter/WARCnet/Maemura\\_Towards\\_an\\_Infrastructural\\_Description.pdf](https://cc.au.dk/fileadmin/dac/Projekter/WARCnet/Maemura_Towards_an_Infrastructural_Description.pdf). [URL Memento: Wayback Machine]
- Maguire, M. (2022). 16 January 1922: The 'Surrender of Dublin Castle' [Web page]. RTÉ: Century Ireland. Retrieved 2022-05-15, from <https://www.rte.ie/centuryireland/index.php/articles/16-january-1922-the-surrender-of-dublin-castle> [URL Memento: Wayback Machine]
- Malone, D. (n.d.). Early Irish Web Stuff [Web page]. TCD Maths. Retrieved 2017-05-16, from <https://www.maths.tcd.ie/~dwmalone/early-web.html>. [URL Memento: Wayback Machine]
- Markwell, J., & Brooks, D. W. (2002). Broken Links: The Ephemeral Nature of Educational WWW Hyperlinks. *Journal of Science Education and Technology*, 11(2), 105–108. DOI: 10.1023/A:1014627511641
- Masanès, J. (2005). Web Archiving Methods and Approaches: A Comparative Study. *Library Trends*, 54(1), 72–90. DOI: 10.1353/lib.2006.0005
- Masanès, J. (2006). Web Archiving: Issues and Methods. In J. Masanès (Ed.), *Web Archiving* (pp. 1–53). Berlin, Heidelberg: Springer-Verlag.
- Matthews, B. (1890) The Evolution of Copyright. *Political Science Quarterly*, 5(4), pp 583-602. DOI: <https://www.jstor.org/stable/2139530>.
- Maurer, Y. (2022, August 10). Investigate holdings of web archives through summaries: Cdx-summarize. Retrieved 2022-08-10, from <https://netpreserveblog.wordpress.com/2022/08/10/investigate-holdings-of-web-archives-through-summaries-cdx-summarize/>. [URL Memento: Wayback Machine]
- Maynooth University. (2016). Maynooth University Research Integrity Policy. Maynooth University. Retrieved 2017-12-02, from [https://www.maynoothuniversity.ie/sites/default/files/assets/document/MU%20Research%20Integrity%20Policy%20September%202016%20\\_2.pdf](https://www.maynoothuniversity.ie/sites/default/files/assets/document/MU%20Research%20Integrity%20Policy%20September%202016%20_2.pdf). [URL Memento: Wayback Machine]
- Maynooth University. (2019). Maynooth University Online Surveys (formerly Bristol Online Survey) User Policy. Maynooth University. Retrieved 2022-02-15, from <https://www.maynoothuniversity.ie/sites/default/files/assets/document/Maynooth%20University%20OnlineSurveys%20User%20Policy%20FINAL.pdf>. [URL Memento: Archive.today]
- Maynooth University. (2020). Maynooth University Research Ethics Policy (Updated March 2020). Maynooth University. Retrieved 2020-11-10, from <https://www.maynoothuniversity.ie/sites/default/files/assets/document//Maynooth%20University%20Research%20Ethics%20Policy%20%28Updated%20March%202020%29.pdf>. [URL Memento: Wayback Machine]
- Maynooth University. (2021). Maynooth University Research Integrity Policy. Maynooth University. Retrieved 2021-10-05, from <https://www.maynoothuniversity.ie/sites/default/files/assets/document//MU%20Research>

- rch%20Integrity%20%20Policy%20V4.0%2026%2004%20%2021\_approved%20by%20Research%20Committee.pdf. [URL Memento: Wayback Machine]
- McDonald, H. (2008, April 2). Irish prime minister Ahern resigns amid financial controversy. *The Guardian*. Retrieved 2021-10-23, from <https://www.theguardian.com/world/2008/apr/02/ireland> [URL Memento: Wayback Machine]
- McGee, H. (2018, December 28). Move to begin in 2019 to release State papers after 20 years. *The Irish Times*. Retrieved 2018-12-28, from <https://www.irishtimes.com/news/politics/move-to-begin-in-2019-to-release-state-papers-after-20-years-1.3742288> [URL Memento: Wayback Machine]
- Meyer, E. T., Thomas, A., & Schroeder, R. (2011). *Web Archives: The Future(s)*. Oxford Internet Institute; International Internet Preservation Consortium. Retrieved 2021-09-10, from [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1830025](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1830025). DOI: 10.2139/ssrn.1830025
- Meyer, E. T., Yasseri, T., Halt, S. A., Josh Cows, Schroeder, R., & Helen Margetts. (2017). Analysing the UK web domain and exploring 15 years of UK universities on the web. In N. Brügger & R. Schroeder (Eds.), *The Web as History: Using Web Archives to Understand the Past and the Present* (Online/pdf, pp. 23–44). London: UCL Press. DOI: 10.14324/111.9781911307563 [URL Memento: Wayback Machine]
- Michel, A., Pranger, J., Geeraert, F., Lieber, S., Mechant, P., Vlassenroot, E., Chambers, S., Birkholz, J., & Messens, F. (2021). WP1 report: An international review of Social Media Archiving initiatives [BESOCIAL – Report WP1]. Retrieved 2022-07-20, from <https://orfeo.belnet.be/handle/internal/7741>. [URL Memento: Wayback Machine]
- Milligan, I. (2019). *History in the Age of Abundance? How the Web is Transforming Historical Research*. Canada: McGill-Queen's University Press.
- Milligan, I. (2015, July 14). Web Archive Legal Deposit: A Double-Edged Sword. Ian Milligan: Digital History, Web Archives, and Contemporary History. Retrieved 2022-01-24, from <https://ianmilli.wordpress.com/2015/07/14/web-archive-legal-deposit-a-double-edged-sword/>. [URL Memento: Wayback Machine]
- Millward, G. (2015). Digital barriers and the accessible web: Disabled people, information and the internet (Big UK Domain Data for the Arts and Humanities) [Case study]. School of Advanced Study, University of London. Retrieved 2022-01-25, from <http://sas-space.sas.ac.uk/6104/> [URL Memento: Wayback Machine]
- MirrorWeb. (n.d.). SEC 17a-4 (Electrolyte) [Web Page]. MirrorWeb. Retrieved 2022-08-10, from <https://www.mirrorweb.com/solutions/sec-17a-4>. [URL Memento: Wayback Machine]
- Mirtaheri, S. M., Dinçtürk, M. E., Hooshmand, S., Bochmann, G. V., Jourdan, G.-V., & Onut, I. V. (2013). A brief history of web crawlers. *Proceedings of the 2013 Conference of the Center for Advanced Studies on Collaborative Research*, 40–54. DOI: 10.5555/2555523.2555529; Preprint: <http://arxiv.org/abs/1405.0749>
- Mohr, G., Michael F. Stack, Igor Ranitovic, Daniel Avery, & Michele Kimpton. (2004). Introduction to Heritrix, an archival quality web crawler. *Proceedings of the 4th International Web Archiving Workshop (IWA'04)*, Bath, UK, July 2004, pp. 1–15. Retrieved 2014-07-18, from

- <https://www.bibsonomy.org/bibtex/209d70d4ea1810fe89522755a0982169f/jaeschke>.  
[URL Memento: Wayback Machine]
- Moiraghi, E. (2018). *Le projet Corpus et ses publics potentiels*. [Research Report] Une étude prospective sur les besoins et les attentes des futurs usagers. Bibliothèque nationale de France. Retrieved 2021-04-12, from <https://hal-bnf.archives-ouvertes.fr/hal-01739730>.  
[URL Memento: Wayback Machine]
- Morris, J. W. (2019). Hearing the Past: The Sonic Web from MIDI to Music Streaming. In N. Brügger & I. Milligan (Eds.), *The SAGE Handbook of Web History* (pp. 491–504). London: SAGE Publications.
- Muir, A. (2005). Legal deposit of digital publications [PhD Dissertation, Loughborough University]. Retrieved 2022-02-25, from [https://repository.lboro.ac.uk/articles/thesis/Legal\\_deposit\\_of\\_digital\\_publications/9415538/1](https://repository.lboro.ac.uk/articles/thesis/Legal_deposit_of_digital_publications/9415538/1). uk.bl.ethos.444005
- Murray, K. R., & Hsieh, I., K. (2008). Archiving Web-published materials: A needs assessment of librarians, researchers, and content providers. *Government Information Quarterly*, 25(1), 66–89. DOI: 10.1016/j.giq.2007.04.005
- Murchan, R. (2020a-09). PRONI Web Archive: A Collaborative Approach [Conference presentation (video; abstract)]. Engaging with Web Archives: ‘Opportunities, Challenges and Potentialities’, (#EWAVirtual), 21-22 September 2020, Maynooth University Arts and Humanities Institute, Co. Kildare, Ireland, [online].  
[https://www.youtube.com/watch?v=OT\\_DXyhUCD4&feature=youtu.be](https://www.youtube.com/watch?v=OT_DXyhUCD4&feature=youtu.be) [YouTube] DOI: 10.5281/zenodo.4058013
- Murchan, R. (2020b-10-28). PRONI Web Archive: A Collaborative Approach (Blog) [Blog post]. UK Web Archive Blog. Retrieved 2022-01-08, from <https://blogs.bl.uk/webarchive/2020/10/proni-web-archive-a-collaborative-approach.html> [URL Memento: Wayback Machine]
- Musso, M., & Merletti, F. (2016). This is the future: A reconstruction of the UK business web space (1996–2001). *New Media & Society*, 18(7), 1120–1142. DOI: 10.1177/1461444816643791
- Nanni, F. (2017). Reconstructing a website’s lost past Methodological issues concerning the history of Unibo.it. *Digital Humanities Quarterly*, 011(2). Retrieved 2018-08-06, from <http://www.digitalhumanities.org/dhq/vol/11/2/000292/000292.html>. [URL Memento: Wayback Machine]
- National Archives of Ireland. (n.d.). News: Shaping Our Future in the Information Age – National Archives Strategic Plan 2021-2025 [Web page]. National Archives of Ireland (NAI). Retrieved 2022-05-20, from <https://www.nationalarchives.ie/news/shaping-our-future-in-the-information-age-national-archives-strategic-plan-2021-2025/>. [URL Memento: Wayback Machine]
- National Archives of Ireland. (n.d.). Public Record Office of Ireland: The Story of a Building—2022 Commemoration Programme [Web page]. National Archives of Ireland (NAI). Retrieved 2022-07-05, from <https://www.nationalarchives.ie/2021commemorationprogramme/public-record-office-of-ireland-the-story-of-a-building-exhibition/>. [URL Memento: Wayback Machine]

National Archives of Ireland. *Report of the Director of the National Archives for 2012*. National Archives of Ireland (NAI). Retrieved 2019-12-10, from <https://www.nationalarchives.ie/wp-content/uploads/2019/03/DirectorsReport2012.pdf>. [URL Memento: Wayback Machine]

National Archives of Ireland. *Report of the Director of the National Archives for 2013*. National Archives of Ireland (NAI). Retrieved 2019-12-10, from <https://www.nationalarchives.ie/wp-content/uploads/2019/03/DirectorsReport2014.pdf>. [URL Memento: Wayback Machine]

National Archives of Ireland. *Report of the Director of the National Archives for 2014*. National Archives of Ireland (NAI). Retrieved 2019-12-10, from <https://www.nationalarchives.ie/wp-content/uploads/2019/03/DirectorsReport2014.pdf>. [URL Memento: Wayback Machine]

National Archives of Ireland. *Report of the Director of the National Archives for 2015*. National Archives of Ireland (NAI). Retrieved 2019-12-10, from <https://www.nationalarchives.ie/wp-content/uploads/2019/03/DirectorsReport2015.pdf>. [URL Memento: Wayback Machine]

National Archives of Ireland. *Report of the Director of the National Archives for 2016*. National Archives of Ireland (NAI). Retrieved 2019-12-10, from <https://www.nationalarchives.ie/wp-content/uploads/2019/03/National-Interactive-Archives-Report.pdf>. [URL Memento: Wayback Machine]

National Archives of Ireland. *Report of the Director of the National Archives for 2017*. National Archives of Ireland (NAI). Retrieved 2019-12-10, from [https://www.nationalarchives.ie/wp-content/uploads/2019/06/18-092-National-Archives-Report-08.01.19\\_V08-002.pdf](https://www.nationalarchives.ie/wp-content/uploads/2019/06/18-092-National-Archives-Report-08.01.19_V08-002.pdf). [URL Memento: Wayback Machine]

National Archives of Ireland. *Report of the Director of the National Archives for 2018*. National Archives of Ireland (NAI). Retrieved 2019-12-10, from [https://www.nationalarchives.ie/wp-content/uploads/2021/05/NA\\_2018-Annual-Report-redux.pdf](https://www.nationalarchives.ie/wp-content/uploads/2021/05/NA_2018-Annual-Report-redux.pdf). [URL Memento: Wayback Machine]

National Archives of Ireland. *Report of the Director of the National Archives for 2019*. National Archives of Ireland (NAI). Retrieved 2021-09-15, from <https://www.nationalarchives.ie/wp-content/uploads/2021/07/21.03.15-NA-AR-2019-Redux.pdf>. [URL Memento: Wayback Machine]

National Archives of Ireland. *Report of the Director of the National Archives for 2020*. National Archives of Ireland. Retrieved 2022-03-02, from [https://www.nationalarchives.ie/wp-content/uploads/2022/01/21.10.01-NA-AR-2020-English-and-Gaeilge-V001\\_Directors-Exp-compressed\\_2.pdf](https://www.nationalarchives.ie/wp-content/uploads/2022/01/21.10.01-NA-AR-2020-English-and-Gaeilge-V001_Directors-Exp-compressed_2.pdf). [URL Memento: Wayback Machine]

National Archives of Ireland. (2021-03-11). *Shaping our future in the Information Age: The National Archives 2021—2025 Strategic Plan*. National Archives of Ireland. Retrieved 2022-03-02, from [https://www.nationalarchives.ie/wp-content/uploads/2021/03/NA-Strategic-Plan-2021-25\\_FINAL\\_RSZ.pdf](https://www.nationalarchives.ie/wp-content/uploads/2021/03/NA-Strategic-Plan-2021-25_FINAL_RSZ.pdf). [URL Memento: Wayback Machine]

National Digital Stewardship Alliance. (2022, April 14). About the NDSA [Web page]. National Digital Stewardship Alliance - Digital Library Federation. Retrieved 2022-04-16, from <http://ndsa.org/about/>. [URL Memento: Wayback Machine]



- National Digital Stewardship Alliance. (n.d.). National Digital Stewardship Alliance [Web page]. National Digital Stewardship Alliance - Digital Library Federation. Retrieved 2021-11-24, from <http://ndsa.org/>. [URL Memento: Wayback Machine]
- National Digital Stewardship Alliance Content Working Group. (2012). *Web Archiving Survey Report*. [NDSA Report]. USA: National Digital Stewardship Alliance (NDSA). Retrieved 2022-03-19, from <https://osf.io/na24q/>. [URL Memento: Archive.today]
- National Library of Ireland. (2009). *Collection Development Policy 2009-2011*. [Web archive]. Dublin, Ireland: National Library of Ireland (NLI). Retrieved from <https://web.archive.org/web/20211125013643/https://www.nli.ie/GetAttachment.aspx?id=be415641-55bb-467c-aa45-50da5c389f78>. [Web Archive: Wayback Machine; source URL: <https://www.nli.ie/GetAttachment.aspx?id=be415641-55bb-467c-aa45-50da5c389f78>; Timestamp: 2021-11-25 01:36:43]
- National Library of Ireland. (2022). *Collection Development Policy 2022-2026*. Dublin, Ireland: National Library of Ireland (NLI). Retrieved from [https://www.nli.ie/sites/default/files/2022-/nli\\_collections\\_revisedpolicy\\_2022.pdf](https://www.nli.ie/sites/default/files/2022-/nli_collections_revisedpolicy_2022.pdf). [URL Memento: Wayback Machine]
- National Library of Ireland. (2011). *NLI Annual Report, 2011*. [Web archive]. Dublin, Ireland: National Library of Ireland (NLI). Retrieved from <https://web.archive.org/web/20211125013643/https://www.nli.ie/GetAttachment.aspx?id=be415641-55bb-467c-aa45-50da5c389f78>. [Web Archive: Wayback Machine; source URL: <https://www.nli.ie/GetAttachment.aspx?id=be415641-55bb-467c-aa45-50da5c389f78>; Timestamp: 2021-11-25 01:36:43]
- National Library of Ireland. (2012). *NLI Annual Report, 2012*. [Web archive]. Dublin, Ireland: National Library of Ireland (NLI). Retrieved from <https://web.archive.org/web/20211125013640/https://www.nli.ie/GetAttachment.aspx?id=f095220a-4e33-47cf-b64f-28850e05955d>. [Web Archive: Wayback Machine; source URL: <https://www.nli.ie/GetAttachment.aspx?id=f095220a-4e33-47cf-b64f-28850e05955d>; Timestamp: 2021-11-25 01:36:40]
- National Library of Ireland. (2014). *Legal Deposit 2014*. [Web Archive]. Dublin, Ireland: National Library of Ireland (NLI). Retrieved from <https://web.archive.org/web/20211125015832/https://www.nli.ie/GetAttachment.aspx?id=e2a3e901-8860-4473-9d2b-77741b45f78f>. [Web archive: Wayback Machine; source URL: <https://www.nli.ie/GetAttachment.aspx?id=e2a3e901-8860-4473-9d2b-77741b45f78f>; Timestamp: 2021-11-25 01:58:32]
- National Library of Ireland. (2022). About the Library. [Archived web page]. National Library of Ireland (NLI). Retrieved from <https://web.archive.org/web/20220619021147/https://www.nli.ie/en/intro/about-the-library.aspx>. [Web archive: Wayback Machine; source URL: <https://www.nli.ie/en/intro/about-the-library.aspx>; Timestamp: 2022-06-19 02:11:47]
- National Library of Ireland. (2022). History of the Library. [Archived web page]. National Library of Ireland (NLI). Retrieved from <https://web.archive.org/web/20221116002213/https://www.nli.ie/en/history-of-the->

- library.aspx [Web Archive: Wayback Machine; source URL: <https://www.nli.ie/en/history-of-the-library.aspx>; Timestamp: 2022-11-16 00:22:13]
- National Library of Ireland. (2022). History of the Library: 1877 to 1926. [Archived web page]. National Library of Ireland (NLI). <https://web.archive.org/web/20221116002213/https://www.nli.ie/en/history-of-the-library-1877-1926.aspx>. [Web archive: Wayback Machine; source URL: <https://www.nli.ie/en/history-of-the-library-1877-1926.aspx>; Timestamp: 2022-11-16 00:22:13]
- National Library of Ireland. (2022). History of the Library: 1927 to Present. [Archived web page]. National Library of Ireland (NLI). <https://web.archive.org/web/20220817171142/https://www.nli.ie/en/history-of-the-library-1927-to-present.aspx>. [Web archive: Wayback Machine; source URL: <https://www.nli.ie/en/history-of-the-library-1927-to-present.aspx>; Timestamp: 2022-08-17 17:11:42]
- National Library of Ireland. (2022). History of the Library: Origins. [Archived web page]. National Library of Ireland (NLI). <https://web.archive.org/web/20220817170613/https://www.nli.ie/en/history-of-the-library-origins.aspx>. [Web archive: Wayback Machine]
- National Library of Ireland. (2022). Irish Domain Web Archive. National Library of Ireland (NLI). Retrieved 2022-09-26, from <https://www.nli.ie/en/irish-domain-web-archive.aspx> [URL Memento: Wayback Machine]
- National Library of Ireland. (2022). Selective Web Archive Collections. National Library of Ireland (NLI). Retrieved 2022-09-29, from <https://www.nli.ie/en/udlist/web-archive-collections.aspx> [URL Memento: Wayback Machine]
- National Library of Ireland. (2022). *NLI 2022—2026 Strategy*. Dublin, Ireland: National Library of Ireland (NLI). Retrieved from <https://www.nli.ie/sites/default/files/2022-11/nlistrategyenglishweb2022.pdf>. [URL Memento: Wayback Machine]
- Newing, C., & Clegg, P. (2021, February 9). Making the UK Government Social Media Archive even better [Blog post]. IIPC Blog. Retrieved 2021-11-04, from <https://netpreserveblog.wordpress.com/2021/02/09/making-the-uk-government-social-media-archive-even-better/>. [URL Memento: Wayback Machine]
- Nidirect. (2015, December 9). About the PRONI Web Archive. Nidirect Government Services. Retrieved 2021-12-17, from <https://www.nidirect.gov.uk/articles/about-proni-web-archive> [URL Memento: Wayback Machine]
- Nidirect. (2016, March 4). Public Record Office of Northern Ireland (PRONI). Nidirect Government Services. Retrieved 2022-09-01, from <https://www.nidirect.gov.uk/campaigns/public-record-office-northern-ireland-proni> [URL Memento: Wayback Machine]
- Nidirect. (n.d.). Getting to PRONI and opening hours. Nidirect Government Services. Retrieved 2021-06-16, from <https://www.nidirect.gov.uk/articles/getting-proni-and-opening-hours> [URL Memento: Wayback Machine]

- Nielsen, J. (2016). *Using Web Archives in Research—An Introduction* (Online/pdf). Aarhus, Denmark: NetLab. Retrieved 2022-06-22, from [https://dighumlab.org/wp-content/uploads/2017/06/Nielsen\\_Using\\_Web\\_Archives\\_in\\_Research.pdf](https://dighumlab.org/wp-content/uploads/2017/06/Nielsen_Using_Web_Archives_in_Research.pdf). [URL Memento: Archive.today]
- Niu, J. (2012). An Overview of Web Archiving. *D-Lib Magazine*, 18(3/4). DOI: 10.1045/march2012-niu1 [URL Memento: Wayback Machine]
- Ntoulas, A., Cho, J., & Olston, C. (2004). What's New on the Web?: The Evolution of the Web from a Search Engine Perspective. Proceedings of the 13th International Conference on World Wide Web (WWW '04), 1–12. ACM Digital Library. DOI: 10.1145/988672.988674
- O'Connell, H. (2022, March 13). Sinn Féin wipes years of media statements from website. *Sunday Independent*. Retrieved 2022-03-13, from <https://www.independent.ie/irish-news/news/sinn-fein-wipes-years-of-media-statements-from-website-41440873.html>. [URL Memento: Wayback Machine]
- O'Dell, E. (2012, February 29). Copyright and Innovation – The CRC Consultation Paper [Blog post]. CEARTA.IE, the Irish for Rights. Retrieved 2020-11-28, from <http://www.cearta.ie/2012/02/copyright-and-innovation-the-crc-consultation-paper/> [URL Memento: Wayback Machine]
- O'Dell, E. (2013, October 29). Modernising Copyright: The Report of the Copyright Review Committee #CRC13 [Blog post]. CEARTA.IE, the Irish for Rights. Retrieved 2022-01-16, from <http://www.cearta.ie/2013/10/modernising-copyright-the-report-of-the-copyright-review-committee/> [URL Memento: Wayback Machine]
- O'Dell, E. (2016a, August 4). The Copyright and Related Rights (Miscellaneous Provisions) Bill 2016 is announced [Blog post]. CEARTA.IE, the Irish for Rights. Retrieved 2022-05-19, from <http://www.cearta.ie/2016/08/the-copyright-and-related-rights-miscellaneous-provisions-bill-2016-is-announced/> [URL Memento: Wayback Machine]
- O'Dell, E. (2016b, November 18). Copyright reform and digital deposit [Blog post]. CEARTA.IE, the Irish for Rights. Retrieved 2020-08-17, from <http://www.cearta.ie/2016/11/copyright-reform-and-digital-deposit/> [URL Memento: Wayback Machine]
- O'Dell, E. (2017a, May 11). Legal deposit of digital publications [Blog post]. CEARTA.IE, the Irish for Rights. Retrieved 2020-08-17, from <http://www.cearta.ie/2017/05/legal-deposit-of-digital-publications/> [URL Memento: Wayback Machine]
- O'Dell, E. (2017b, October 2). The copyright implications of a publicly curated online archive of Oireachtas debates [Blog post]. CEARTA.IE, the Irish for Rights. Retrieved 2022-05-11, from <http://www.cearta.ie/2017/10/the-copyright-implications-of-a-publicly-curated-online-archive-of-oireachtas-debates/> [URL Memento: Wayback Machine]
- O'Dell, E. (2018, October 3). Digital deposit and harvesting the .ie domain [Blog post]. CEARTA.IE, the Irish for Rights. Retrieved 2022-01-19, from <http://www.cearta.ie/2018/10/digital-deposit-and-harvesting-the-ie-domain/> [URL Memento: Wayback Machine]
- O'Dell, E. (2019, May 17). Not archiving the .ie domain, and the death of new politics [Blog post]. CEARTA.IE, the Irish for Rights. Retrieved 2022-01-19, from

- <http://www.cearta.ie/2019/05/not-archiving-the-ie-domain-and-the-death-of-new-politics/> [URL Memento: Wayback Machine]
- O'Dell, E. (2020, May 1). Irish copyright law must enable digital deposit (corrected and updated 11 December 2020) [Blog post]. CEARTA.IE, the Irish for Rights. Retrieved 2020-12-11, from <http://www.cearta.ie/2020/05/irish-copyright-law-must-enable-digital-deposit/> [URL Memento: Wayback Machine]
- Ogden, J. (2021). "Everything on the internet can be saved": Archive Team, Tumblr and the cultural significance of web archiving. *Internet Histories: Digital Technology, Culture and Society*, 0(0), 1–20. DOI: 10.1080/24701475.2021.1985835
- Ogden, J., Halford, S., & Carr, L. (2017). Observing Web Archives: The Case for an Ethnographic Study of Web Archiving. *Proceedings of the 2017 ACM on Web Science Conference*, 299–308. DOI: 10.1145/3091478.3091506
- Ogden, J., Kurzmeier, M., Clavert, F., & Currie, M. (2022). How to Design Web Archives Research. *SAGE Research Methods: Doing Research Online*. DOI: 10.4135/9781529610147
- Ogden, J., & Maemura, E. (2021). 'Go fish': Conceptualising the challenges of engaging national web archives for digital research. *International Journal of Digital Humanities*, 2(1), 43–63. DOI: 10.1007/s42803-021-00032-5
- Old Dominion University, Web Science and Digital Libraries Research Group at Old Dominion University. (n.d.). Web Science and Digital Libraries Research Group (ODU WS-DL) [Blog site]. Web Science and Digital Libraries Research Group. Retrieved 2021-07-29, from <https://ws-dl.blogspot.com/>. [URL Memento: Wayback Machine]
- Partridge, R. C. B. (1938). The History of the Legal Deposit of Books Throughout the British Empire: A Thesis Approved for the Honours Diploma of the Library Association. London: The Library Association. [Internet Archive Open Library]
- Paßmann, J., Helmond, A., & Jansma, R. (2022). From healthy communities to toxic debates: Disqus' changing ideas about comment moderation. *Internet Histories: Digital Technology, Culture and Society*, 0(0), 1–21. DOI: 10.1080/24701475.2022.2105123
- Pennock, M. (2013). *Web-Archiving* [Report]. DPC Technology Watch Report 13. UK: Digital Preservation Coalition in association with Charles Beagrie Ltd. Retrieved 2019-02-08, from <https://www.dpconline.org/docs/technology-watch-reports/865-dpctw13-01-pdf/file>. DOI: 10.7207/twr13-01 [URL Memento: Wayback Machine]
- Pettenati, C. (2002). Electronic publishing at the end of 2001. In M. Barone, E. Borch, J. Huston, C Leroy, P. G. Rancoita, P. Riboni, & R. Ruchti (Eds.), *Astroparticle, Particle, Space Physics, Radiation Interaction, Detectors and Medical Physics Applications: Volume 1* (Vol. 1, pp. 525–533). World Scientific. [https://doi.org/10.1142/9789812776464\\_0076](https://doi.org/10.1142/9789812776464_0076) . Preprint: [https://web.archive.org/web/20170808230408/http://villaolmo.mib.infn.it/Manuscripts/10\\_generalities/pettenati.pdf](https://web.archive.org/web/20170808230408/http://villaolmo.mib.infn.it/Manuscripts/10_generalities/pettenati.pdf).
- Public Record Office of Northern Ireland. (2007). Public Record Office: A Brief History [Document]. Northern Ireland: Public Record Office of Northern Ireland. Retrieved 2016-09-02, from <https://www.nidirect.gov.uk/sites/default/files/publications/general-information-series-the-public-record-office-a-brief-history-2.pdf>. [URL Memento: Wayback Machine]

- Public Record Office of Northern Ireland. (2008). Public Record Office of Northern Ireland (PRONI): Past, Present and Future [Document]. Northern Ireland: Public Record Office of Northern Ireland. Retrieved from [https://web.archive.org/web/20081120000333/http://www.proni.gov.uk/history\\_of\\_proni\\_-\\_past\\_\\_present\\_and\\_future.pdf](https://web.archive.org/web/20081120000333/http://www.proni.gov.uk/history_of_proni_-_past__present_and_future.pdf). [Web archive: Wayback Machine; source URL: [http://www.proni.gov.uk/history\\_of\\_proni\\_-\\_past\\_\\_present\\_and\\_future.pdf](http://www.proni.gov.uk/history_of_proni_-_past__present_and_future.pdf); Timestamp: 2008-11-20 00:03:33]
- Public Record Office of Northern Ireland. (2016, August 19). PRONI Takedown Policy [Document]. Northern Ireland: Public Record Office of Northern Ireland. Retrieved 2020-10-06, from [https://www.nidirect.gov.uk/sites/default/files/publications/PRONI%20Takedown%20Policy\\_0.pdf](https://www.nidirect.gov.uk/sites/default/files/publications/PRONI%20Takedown%20Policy_0.pdf) [URL Memento: Wayback Machine]
- Public Record Office of Northern Ireland. (2018, November). PRONI Web Archiving Strategy [Document]. Northern Ireland: Public Record Office of Northern Ireland. Retrieved 2020-07-30, from [https://www.nidirect.gov.uk/sites/default/files/publications/Web%20Archiving%20Strategy\\_1.pdf](https://www.nidirect.gov.uk/sites/default/files/publications/Web%20Archiving%20Strategy_1.pdf) [URL Memento: Wayback Machine]
- Raffal, H. (2018). Tracing the online development of the Ministry of Defence and Armed Forces through the UK web archive. *Internet Histories: Digital Technology, Culture and Society*, 2(1–2), 156–178. DOI: 10.1080/24701475.2018.1456739
- Ras, M., & van Bussel, S. (2007). *Web Archiving User Survey* [Technical Report]. National Library of the Netherlands (Koninklijke Bibliotheek). Retrieved from [http://web.archive.org/web/20220120040514/https://www.kb.nl/sites/default/files/docs/kb\\_usersurvey\\_webarchive\\_en.pdf](http://web.archive.org/web/20220120040514/https://www.kb.nl/sites/default/files/docs/kb_usersurvey_webarchive_en.pdf). [Web archive: Wayback Machine; source URL: [https://www.kb.nl/sites/default/files/docs/kb\\_usersurvey\\_webarchive\\_en.pdf](https://www.kb.nl/sites/default/files/docs/kb_usersurvey_webarchive_en.pdf); Timestamp: 2022-01-20 04:05:14]
- Regan, J. M. (2016). Kindling the Singing Flame: The Destruction of the Public Record Office (30 June 1922) as a Historical Problem. *The Old Athlone Society*, pp. 107–123. Retrieved 2022-10-31, from [https://www.academia.edu/30977460/Kindling\\_the\\_Singing\\_Flame\\_The\\_Destruction\\_of\\_the\\_Public\\_Record\\_Office\\_30\\_June\\_1922\\_as\\_a\\_Historical\\_Problem](https://www.academia.edu/30977460/Kindling_the_Singing_Flame_The_Destruction_of_the_Public_Record_Office_30_June_1922_as_a_Historical_Problem) [URL Memento: Wayback Machine]
- Recite Me. (n.d.). Recite Me: Choosing an Accessible Font [Web page/pdf]. Recite Me. Retrieved 2021-12-16, from [https://reciteme.com/uploads/articles/accessible\\_fonts\\_guide.pdf](https://reciteme.com/uploads/articles/accessible_fonts_guide.pdf). [URL Memento: Wayback Machine]
- RESAW. (n.d.). About RESAW [Web page]. Research Infrastructure for the Study of Archived Web Materials. Retrieved 2021-03-10, from <http://resaw.eu/about/>. [URL Memento: Wayback Machine]
- Reyes Ayala, B. (2013). Web Archiving Bibliography 2013. University of North Texas Libraries. Retrieved 2020-07-30, from <https://digital.library.unt.edu/ark:/67531/metadc172362/>. [URL Memento: Wayback Machine]
- Riley, H., & Crookston, M. (2015). *Awareness and Use of the New Zealand Web Archive: A survey of New Zealand academics* [Report]. New Zealand: University of Wellington; National

- Library of New Zealand. Retrieved 2018-01-06, from <https://natlib.govt.nz/files/webarchive/nzwebarchive-awarenessanduse.pdf>. [URL Memento: Wayback Machine]
- Rogers, R. (2013). *Digital Methods*. Cambridge, MA, USA; London, UK: MIT Press.
- Rogers, R. (2017). Doing Web history with the Internet Archive: Screencast documentaries. *Internet Histories: Digital Technology, Culture and Society*, 1(1–2), 160–172. DOI: 10.1080/24701475.2017.1307542
- Rogers, R. (2019). Periodising Web Archiving. In N. Brügger & I. Milligan (Eds.), *The SAGE Handbook of Web History* (pp. 42–56). London: SAGE Publications.
- Rosnay, M. D. de, Georges, F., Crosnier, H. L., Merzeau, L., Musiani, F., Paloque-Berges, C., Schafer, V., & Thierry, B. (Eds.) (n.d.). Web90 – Heritage, Memories and History of the Web in the 1990s [Blog site]. Web90 project. Retrieved 2022-07-09, from <https://web90.hypotheses.org/>. [URL Memento: Wayback Machine]
- Ruest, N. (2016, June 18). Web Archives for Historical Research Group [Community group]. Zenodo. Retrieved 2022-09-14, from <https://zenodo.org/communities/wahr/?page=1&size=20>. [URL Memento: Archive.today]
- Ruest, N., Fritz, S., Deschamps, R., Lin, J., & Milligan, I. (2021). From archive to analysis: Accessing web archives at scale through a cloud-based interface. *International Journal of Digital Humanities*, 2(1), 5–24. DOI: 10.1007/s42803-020-00029-6
- Ruest, N., Lin, J., Milligan, I., & Fritz, S. (2020). The Archives Unleashed Project: Technology, Process, and Community to Improve Scholarly Access to Web Archives. *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020*. Association for Computing Machinery, New York, NY, USA, 157–166. DOI: 10.1145/3383583.3398513
- Ryan, M., Keating, D., & Finegan, J. (2022). Managing and accessing web archives: Irish practitioners’ perspectives. *AI & SOCIETY*. DOI: 10.1007/s00146-021-01364-0 [URL Memento: Wayback Machine]
- Schafer, V. (2019). Exploring the ‘French web’ of the 1990s. In N. Brügger & D. Laursen (Eds.), *The Historical Web and Digital Humanities: The Case of National Web Domains* (pp. 145–160). London & New York: Routledge.
- Schafer, V. (2020). From Print to Digital, from Document to Data: Digitalisation at the Publications Office of the European Union. *Open Information Science*, 4(1), 203–216. DOI: 10.1515/opis-2020-0015
- Schafer, V., Musiani, F., & Borelli, M. (2016). Web archiving, governance and STS. *French Journal for Media Research*, no. 6/2016(6), 23. <http://frenchjournalformediaresearch.com/lodel-1.0/main/index.php?id=952> [URL Memento: Wayback Machine]
- Schafer, V., Truc, G., Badouard, R., Castex, L., & Musiani, F. (2019). Paris and Nice terrorist attacks: Exploring Twitter and web archives. *Media, War & Conflict*, 12(2), 153–170. DOI: 10.1177/1750635219839382
- Schafer, V. & Winters, J. (2021). The values of web archives. *International Journal of Digital Humanities*, 2(1), 129–144. DOI: 10.1007/s42803-021-00037-0 [URL Memento: Wayback Machine]

- Schneider, S. M., Foot, K. A., & Wouters, P. (2009). Web archiving as e- research. In N. W. Jankowski (Ed.), *E-Research: Transformation in Scholarly Practice* (pp. 205–221). New York, London: Routledge.
- Schroeder, R. & Brügger, N. (2017). Introduction: The Web as History. In Niels Brügger & R. Schroeder (Eds.), *The Web as History: Using Web Archives to Understand the Past and the Present* (Online/pdf, pp. 1–22). London: UCL Press. DOI: 10.14324/111.9781911307563 [URL Memento: Wayback Machine]
- Sellitto, C. (2004). A study of missing Web-cites in scholarly articles: Towards an evaluation framework. *Journal of Information Science*, 30(6), 484–495. DOI: 10.1177/0165551504047822
- Sellitto, C. (2005). The impact of impermanent Web-located citations: A study of 123 scholarly conference publications. *Journal of the American Society for Information Science and Technology*, 56(7), 695–703. DOI: 10.1002/asi.20159
- Song, S., & JaJa, J. (2008). Archiving Temporal Web Information: Organization of Web Contents for Fast Access and Compact Storage (Technical Report UMIACS-TR-2008-08). Department of Electrical and Computer Engineering Institute for Advanced Computer Studies, University of Maryland. Retrieved 2020-10-28, from <https://drum.lib.umd.edu/handle/1903/7569>. [URL Memento: Wayback Machine]
- Spinellis, D. (2003). The Decay and Failures of Web References. *Communications of the ACM*, 46(1), 71–77. DOI: 10.1145/602421.602422
- Stanford Libraries. (n.d.). Archivability [Web page]. Stanford University. Retrieved 2022-05-02, from <https://library.stanford.edu/projects/web-archiving/archivability> [URL Memento: Wayback Machine]
- Steber, C. (2016). Online Surveys: Data Collection Advantages & Disadvantages [Blog post]. Communications for Research. Retrieved from <https://web.archive.org/web/20201003184529/https://www.cfrinc.net/cfrblog/online-surveys-advantages-disadvantages>. [Web archive: Wayback Machine; source URL: <https://www.cfrinc.net/cfrblog/online-surveys-advantages-disadvantages>; Timestamp: 2020-10-03 18:45:29]
- Stember, M. (1991). Advancing the social sciences through the interdisciplinary enterprise. *The Social Science Journal*, 28(1), 1–14. DOI: 10.1016/0362-3319(91)90040-B
- Sterne, J. (2015+). How the internet came to Ireland. TechArchives; TechArchives Digital Repository. Retrieved 2019-01-13, from <https://techarchives.irish/how-the-internet-came-to-ireland-1987-97/>. [URL Memento: Wayback Machine]
- Sterne, J. (1995, February 27). IT's Monday One Hundred and Forty-Two—27 February 95. *IT's Monday*, 142. TechArchives, Ireland. TechArchives, Ireland - Repository
- Stirling, P., Chevallier, P., & Illien, G. (2012). Web Archives for Researchers: Representations, Expectations and Potential Uses. *D-Lib Magazine*, 18(3/4). DOI: 10.1045/march2012-stirling [URL Memento: Wayback Machine]
- Taylor, A. (2013, June 8). What Is Digital Publishing, And Is It Necessary? [Web page]. Author Mingle. <https://web.archive.org/web/20130712072819/http://authormingle.com/ebooks/what-is-digital-publishing-and-is-it-necessary/> [Web archive: Wayback Machine; source URL: <https://web.archive.org/web/20130712072819/http://authormingle.com/ebooks/what-is-digital-publishing-and-is-it-necessary/>]

- <http://authormingle.com/ebooks/what-is-digital-publishing-and-is-it-necessary/>; Timestamp: 2013-07-12 07:28:19]
- Taylor, C. (2017a, July 20). Ireland's digital content in danger of disappearing, specialist warns. *The Irish Times* [online]. Retrieved 2017-09-22, from <https://www.irishtimes.com/business/technology/ireland-s-digital-content-in-danger-of-disappearing-specialist-warns-1.3157792>. [URL Memento: Wayback Machine]
- Taylor, N. (2017b). Understanding Legal Use Cases for Web Archives [Conference presentation]. International Internet Preservation Coalition General Assembly and Web Archiving Conference, London, June 2016. Retrieved 2021-12-05, from [https://nullhandle.org/pdf/2017-06-16\\_understanding\\_legal\\_use\\_cases\\_for\\_web\\_archives.pdf](https://nullhandle.org/pdf/2017-06-16_understanding_legal_use_cases_for_web_archives.pdf). [URL Memento: Wayback Machine]
- Taylor, N. (2011, June 16). Web and Twitter Archiving at the Library of Congress [Presentation slides]. Web Archive Globalization Workshop June 16, 2011, Library of Congress, USA. Retrieved 2019-06-06, from <http://eventsarchive.org/sites/default/files/webandtwitterarchivingatthelibraryofcongress-110614202159-phpapp01.pdf>. [URL Memento: Wayback Machine]
- The Archives Unleashed Project. (n.d.). The Archives Unleashed Project [Website]. The Archives Unleashed Project. Retrieved 2021-11-23, from <https://archivesunleashed.org>. [URL Memento: Wayback Machine]
- The Archives Unleashed Project. (n.d.). Archives Unleashed Cohorts (2022-2023) [Web page]. The Archives Unleashed Project. Retrieved 2022-02-17, from <https://archivesunleashed.org/cohorts2022-2023/>. [URL Memento: Wayback Machine]
- The Bodleian Libraries. (n.d.). Legal deposit [Web page]. The Bodleian Libraries, University of Oxford. Retrieved 2022-02-03, from <https://www.bodleian.ox.ac.uk/collections-and-resources/legal-deposit>. [URL Memento: Wayback Machine]
- The Irish News. (2022, March 14). Sinn Féin removes thousands of media statements from its website. *The Irish News*. Retrieved 2022-03-14, from <https://www.irishnews.com/news/northernirelandnews/2022/03/14/news/sinn-fe-in-removes-thousands-of-media-statements-from-its-website-2613771/>. [URL Memento: Wayback Machine]
- The National Archives. (n.d.). The EEC and Britain's late entry [Web page/catalogue]. The National Archives. Retrieved 2022-08-08, from <https://www.nationalarchives.gov.uk/cabinetpapers/themes/eec-britains-late-entry.htm> [URL Memento: Wayback Machine]
- The National Archives. (n.d.). The National Archives – Home [Website]. The National Archives. Retrieved 2021-10-07, from <https://www.nationalarchives.gov.uk>. [URL Memento: Wayback Machine]
- Thomas, A., Meyer, E. T., Dougherty, M., van den Heuvel, C., Madsen, C. M., & Wyatt, S. (2010). *Researcher Engagement with Web Archives: Challenges and Opportunities for Investment*. Joint Information Systems Committee Report, August 2010. UK: JISC.



- Retrieved 2020-04-14, from <https://papers.ssrn.com/abstract=1715000>. [URL Memento: Wayback Machine]
- Thompson, D. (2008). Archiving Web Resources. In S. Ross & M. Day (Eds.), DCC Digital Curation Manual. HATII, University of Glasgow; University of Edinburgh; UKOLN, University of Bath; Council for the Central Laboratory of the Research Councils; Digital Curation Centre. Retrieved 2021-06-04, from <https://www.dcc.ac.uk/sites/default/files/documents/resource/curation-manual/chapters/archiving-web-resources/archiving-web-resources.pdf>. [URL Memento: Wayback Machine]
- Thomson, S. D. (2016). Preserving Social Media (DPC Technology Watch Reports). Digital Preservation Coalition. Retrieved 2019-06-07, from <https://www.dpconline.org/docs/technology-watch-reports/1486-twr16-01/file> [URL Memento: Wayback Machine]
- Trinity College Dublin. (n.d). Trinity College Dublin, the University of Dublin, Ireland [Website]. Trinity College Dublin. Retrieved 2021-12-20, from <https://www.tcd.ie>. [URL Memento: Wayback Machine]
- Truman, G. (2016). *Web Archiving Environmental Scan* [Report]. Harvard Library Report, 2016. USA: Harvard Library. Retrieved 2021-02-24, from <https://dash.harvard.edu/handle/1/25658314>. [URL Memento: Wayback Machine]
- Truter, V. (2021). Research Data Management and Sharing Practices of Researchers in Web Archive Studies [Conference presentation & abstract]. *Engaging with Web Archives 4 Digital Humanities (#EWA4DH)*, Maynooth University Arts and Humanities Institute, Co. Kildare, Ireland, 21 September 2021. Retrieved 2022-05-07, from <https://ewaconference.com/ewa4dh-2021/ewa4dh-programme/>. [URL Memento: Wayback Machine]
- UCD Library. (n.d.). LibGuides: Irish Poetry Reading Archive: Introduction & News [Web page]. UCD Library. Retrieved 2022-10-31, from <https://libguides.ucd.ie/ipra/news> [URL Memento: Archive.today]
- UK Web Archive. (2009, archived). F.A.Q (archived) [Archived web page]. UK Web Archive. <https://www.webarchive.org.uk/wayback/en/archive/20091130210036/http://www.webarchive.org.uk/ukwa/info/faq> [Web archive: UK Web Archive; source URL: <http://www.webarchive.org.uk/ukwa/info/faq>; Timestamp: 2009-11-30 21:00:36]
- UK Web Archive. (2020, March 2). 15 Years of the UK Web Archive—The Early Years. Retrieved 2020-03-02, from <https://blogs.bl.uk/webarchive/2020/03/15-years-of-the-uk-web-archive.html> [URL Memento: Wayback Machine]
- UK Web Archive. (n.d.). Frequently Asked Questions [Web page]. UK Web Archive. Retrieved 2022-06-19, from <https://www.webarchive.org.uk/en/ukwa/info/faq/> [URL Memento: Wayback Machine]
- UK Web Archive. (2018, February 1). A New Playback Tool for the UK Web Archive [Blog post]. UK Web Archive Blog. Retrieved 2021-05-12, from <https://blogs.bl.uk/webarchive/2018/02/index.html>. [URL Memento: UK Web Archive]

- UNESCO. (2003). UNESCO Charter on the Preservation of the Digital Heritage. United Nations Educational, Scientific and Cultural Organization. Retrieved from <https://unesdoc.unesco.org/ark:/48223/pf0000229034>. Document code: IFAP-2003/COUNCIL.II/4
- UNESCO & IFLA-PAC. (n.d.). Safeguarding our documentary heritage [Web page]. Webworld.Unesco.Org. Retrieved from <https://web.archive.org/web/20010429033355/http://webworld.unesco.org/safeguarding/en/index.html> [Web archive: Wayback Machine; source: URL: <http://webworld.unesco.org/safeguarding/en/index.html>; Timestamp: 2001-04-29 03:33:55]
- University College Dublin. (2021, December 20). UCD appoints Dr Sandra Collins as Librarian. Retrieved 2022-10-01, from <https://www.ucd.ie/newsandopinion/news/2021/december/20/ucdappointsdrsandracoillinsaslibrarian/> [URL Memento: Wayback Machine]
- Vlassenroot, E., Chambers, S., Di Pretoro, E., Geeraert, F., Haesendonck, G., Michel, A., & Mechant, P. (2019). Web archives as a data resource for digital scholars. *International Journal of Digital Humanities*, 1(1), 85–111. DOI: 10.1007/s42803-019-00007-7
- Vlassenroot, E., Chambers, S., Lieber, S., Michel, A., Geeraert, F., Pranger, J., Birkholz, J., & Mechant, P. (2021). Web-archiving and social media: An exploratory analysis. *International Journal of Digital Humanities*, 2(1), 107–128. DOI: 10.1007/s42803-021-00036-1 [URL Memento: Wayback Machine]
- Waniata, R. (2018, February 7). The Life and Times of the Late, Great CD: Remembering the rise (and final fall) of the late, great Compact Disc. *Digital Trends*. Retrieved 2022-06-19, from <https://www.digitaltrends.com/music/the-history-of-the-cds-rise-and-fall/> . [URL Memento: Wayback Machine]
- WARCnet. (n.d.). About WARCnet [Web page]. Aarhus University. Retrieved 2021-04-23, from <https://cc.au.dk/en/warcnet/about/>. [URL Memento: Wayback Machine]
- Waters, D., & Garrett, J. (1996). Preserving Digital Information. Report of the Task Force on Archiving of Digital Information. The Commission on Preservation and Access and The Research Libraries Group. Retrieved 2021-05-07, from <https://www.clir.org/pubs/reports/pub63/>. [URL Memento: Wayback Machine]
- Weber, M. (2019). Browser and Browser Wars. In N. Brügger & I. Milligan (Eds.), *The SAGE Handbook of Web History* (pp. 270–298). London: SAGE Publications.
- Weber, M. S., & Napoli, P. M. (2018). Journalism History, Web Archives, and New Methods for Understanding the Evolution of Digital Journalism. *Digital Journalism*, 6(9), 1186–1205. DOI: 10.1080/21670811.2018.1510293
- Webster, P. (2017a). Religious discourse in the archived web: Rowan Williams, Archbishop of Canterbury, and the sharia law controversy of 2008. In N. Brügger & R. Schroeder (Eds.), In Niels Brügger & R. Schroeder (Eds.), *The Web as History: Using Web Archives to Understand the Past and the Present* (Online/pdf, pp. 1–22). London: UCL Press. DOI: 10.14324/111.9781911307563 [URL Memento: Wayback Machine]

- Webster, P. (2017b). Users, technologies, organisations: Towards a cultural history of world web archiving. In N. Brügger (Ed.), *Web 25: Histories from the first 25 Years of the World Wide Web* (pp. 175–190). New York: Peter Lang.
- Webster, P. (2012, February 2). The AADDA project [Education; Research; Libraries; Academic blog]. AADDA Blog. Retrieved 2020-10-20, from <http://domaindarkarchive.blogspot.com/2012/02/>. [URL Memento: Wayback Machine]
- Webster, P. (2019). Understanding the limitations of the ccTLD as a proxy for the national web: lessons from cross-border religion in the northern Irish web sphere. In N. Brügger & D. Laursen (Eds.), *The Historical Web and Digital Humanities: The Case of National Web Domains* (pp. 110–123). London & New York: Routledge.
- Webster, P. (2020). How Researchers Use the Archived Web. DPC Technology Watch Guidance Note. UK: Digital Preservation Coalition. Retrieved 2021-12-07, from <https://www.dpconline.org/docs/technology-watch-reports/2263-twgn-20-01-how-researchers-use-the-archived-web-webster/file>. DOI 10.7207/twgn20-01 [URL Memento: Wayback Machine]
- Weigle, M. C. (2018, September 19). On the Importance of Web Archiving. Items: Insights from the Social Sciences. Retrieved 2021-03-10, from <https://items.ssrc.org/parameters/on-the-importance-of-web-archiving/>
- Weisbard, P. H. (2011). Oldies but Goodies: Archiving Web- Based Information. *Feminist Collections: A Quarterly of Women's Studies Resources*, 32(2), 14–20. Gale Document Number: GALE|A339529981.
- Weiss, R. (2003, November 24). On the Web, Research Work Proves Ephemeral. *Washington Post* [online] Retrieved 2018-04-29, from <https://www.washingtonpost.com/archive/politics/2003/11/24/on-the-web-research-work-proves-ephemeral/959c882f-9ad0-4b36-88cd-fb7411db118d/> [URL Memento: Wayback Machine]
- Wellcome Trust Library. (2008, September 22). Web archiving: A feasibility study for JISC and the Wellcome Trust [Archived web page]. Wellcome Trust Library. Retrieved from <https://www.webarchive.org.uk/wayback/archive/20080922220153/http://library.wellcome.ac.uk/node228.html> [Web archive: UK Web Archive; source URL: <http://library.wellcome.ac.uk/node228.html>; Timestamp: 2008-09-22 22:01:53]
- Wikipedia. (2002+). Information science (also known as information studies). Wikipedia. Retrieved 2022-02-02, from [https://en.wikipedia.org/w/index.php?title=Information\\_science&oldid=1040622798](https://en.wikipedia.org/w/index.php?title=Information_science&oldid=1040622798). [URL Memento: Archive.today]
- Wikipedia. (2005+). Department of Enterprise, Trade and Employment. Wikipedia. Retrieved 2022-07-07, from [https://en.wikipedia.org/wiki/Department\\_of\\_Enterprise,\\_Trade\\_and\\_Employment](https://en.wikipedia.org/wiki/Department_of_Enterprise,_Trade_and_Employment) [URL Memento: Wayback Machine]
- Wikipedia. (2007+). Department of Tourism, Culture, Arts, Gaeltacht, Sport and Media. Wikipedia. Retrieved 2023-01-05, from [https://en.wikipedia.org/w/index.php?title=Department\\_of\\_Tourism,\\_Culture,\\_Arts,\\_Gaeltacht,\\_Sport\\_and\\_Media&oldid=1130593170](https://en.wikipedia.org/w/index.php?title=Department_of_Tourism,_Culture,_Arts,_Gaeltacht,_Sport_and_Media&oldid=1130593170). [URL Memento: Archive.today]

- Wikipedia. (2009+). Irish Manuscripts Commission. Wikipedia. Retrieved 2023-02-11, from [https://en.wikipedia.org/wiki/Irish\\_Manuscripts\\_Commission](https://en.wikipedia.org/wiki/Irish_Manuscripts_Commission) [URL Memento: Wayback Machine]
- Wikipedia. (2011+). List of Web archiving initiatives. Wikipedia. Retrieved 2021-09-18, from [https://en.wikipedia.org/wiki/List\\_of\\_Web\\_archiving\\_initiatives](https://en.wikipedia.org/wiki/List_of_Web_archiving_initiatives). [URL Memento: Wayback Machine]
- Wikipedia. (2012+). Internet Memory Foundation. Wikipedia. Retrieved 2022-06-27, from [https://en.wikipedia.org/wiki/Internet\\_Memory\\_Foundation](https://en.wikipedia.org/wiki/Internet_Memory_Foundation). [URL Memento: Wayback Machine]
- Winters, J. (2017). Breaking in to the mainstream: Demonstrating the value of internet (and web) histories. *Internet Histories: Digital Technology, Culture and Society*, 1(1–2), 173–179. DOI: 10.1080/24701475.2017.1305713
- Winters, J. (2018). Digital History. In M. Tamm & P. Burke (Eds.), *Debating New Approaches to History* (pp. 277–300). London: Bloomsbury Publishing.
- Winters, J. (2019). Negotiating the archives of UK web space. In N. Brügger & D. Larsen (Eds.), *The Historical Web and Digital Humanities: The Case of National Web Domains* (pp. 75–88). London & New York: Routledge.
- Winters, J. (2020a). Giving with one Click, Taking with the Other: Electronic Legal Deposit, Web Archives and Researcher Access. In M. Terras & P. Gooding (Eds.), *Electronic Legal Deposit: Shaping the Library Collections of the Future* (pp. 159–178). London: Facet Publishing. DOI: 10.29085/9781783303786.010
- Winters, J. (2020b). Web archives as sites of collaboration [Conference keynote presentation (video)]. *Engaging with Web Archives: ‘Opportunities, Challenges and Potentialities’, (#EWAVirtual), Maynooth University Arts and Humanities Institute, Co. Kildare, Ireland, [online], 21-22 September 2022*. Retrieved 2021-09-18, from <https://www.youtube.com/watch?v=c5JYCfnLJ-c>
- Winters, J. (2018). Digital History. In M. Tamm & P. Burke (Eds.), *Debating New Approaches to History* (pp. 277–300). London: Bloomsbury Publishing.
- Winters, J., & Prescott, A. (2019). Negotiating the born-digital: A problem of search. *Archives and Manuscripts*, 47(3), 391–403. DOI: 10.1080/01576895.2019.1640753
- Wren, J. D. (2008). URL decay in MEDLINE—a 4-year follow-up study. *Bioinformatics*, 24(11), 1381–1385. DOI: 10.1093/bioinformatics/btn127
- Wright, R. (1997, May 19). Tim Berners-Lee: The Man Who Invented the Internet [Magazine online]. *Time*, 149(20), p. 1/7. Retrieved 2017-09-21, from <http://content.time.com/time/subscriber/article/0,33009,986354,00.html>. [URL Memento: Wayback Machine]
- Xie, Z., Klein, M., & Fox, E. A. (2020). Web Archiving and Digital Libraries. *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020*, 583–584. DOI: 10.1145/3383583.3398509
- Yusuf, K. F. (2013). The Role of Archives in National Development: National Archives of Nigeria in Perspective. *International Journal of Economic Development Research and Investment*, 4(2), 19–29. Retrieved from

- [http://icidr.org/ijedri\\_vol4\\_no2\\_dec2013/The\\_Role\\_of\\_Archives\\_in\\_National\\_Development\\_Nigeria%20National\\_Archives\\_in\\_Perspective.pdf](http://icidr.org/ijedri_vol4_no2_dec2013/The_Role_of_Archives_in_National_Development_Nigeria%20National_Archives_in_Perspective.pdf) [URL Memento: Archive.today]
- Zhou, K., Grover, C., Klein, M., & Tobin, R. (2015). No More 404s: Predicting Referenced Link Rot in Scholarly Articles for Pro-Active Archiving. *Proceedings of the 15th ACM/IEEE-CS Joint Conference on Digital Libraries*, 233–236. DOI: 10.1145/2756406.2756940
- Zierau, E. M.-B. O. (2019). A Persistent Web IDentifier (PWID) URN Namespace. *Internet Engineering Task Force*. Retrieved 2022-10-31, from, <https://datatracker.ietf.org/doc/draft-pwid-urn-specification-07>. Internet Draft: draft-pwid-urn-specification-07 [URL Memento: Archive. today]
- Zierau, E., Nyvang, C., & Kromann, T. H. (2016). Persistent Web References - Best Practices and New Suggestions. *Proceedings of the 13th International Conference on Digital Preservation (iPRES 2016), Bern, 3-6 October 2016*, pp. 237–246. Retrieved 2022-09-14, from <https://fedora.phaidra.univie.ac.at/fedora/objects/o:502767/methods/bdef:Content/download>.
- Zittrain, J., Albert, K., & Lessig, L. (2014). Perma: Scoping and Addressing the Problem of Link and Reference Rot in Legal Citations. *Legal Information Management*, 14(2), 88-99. DOI: 10.1017/S1472669614000255

## Providers & Services

- Ó Briain, E., & Foster, J. (2020). #retweetthe8th: Twitter dataset from the 2018 Referendum to repeal the 8th Amendment of the Constitution of Ireland. *Zenodo*. DOI: 10.5281/zenodo.3842013
- Archive-It (n.d.). Archive-It—Web Archiving Services for Libraries and Archives [Website]. Archive-It (Internet Archive). Retrieved 2021-03-16, from <https://archive-it.org>. [URL Memento: Wayback Machine]
- archive.today. (n.d.). archive.today: Webpage capture [Website]. archive.today. Retrieved 2021-03- from <https://archive.ph>. [URL Memento: Wayback Machine]
- Arquivo.pt. (n.d.). Arquivo.pt—Pesquise páginas do passado! [Website]. Arquivo.pt. Retrieved 2021-03-23, from <https://arquivo.pt>. [URL Memento: Arquivo.pt]
- Arvidson, A., & Lettenström, F. (1998). The Kulturarw Project—The Swedish Royal Web Archive. *The Electronic Library*, 16(2), 105–108. DOI: 10.1108/eb045623
- Austrian National Library. (n.d.). Webarchiv Österreich [Website]. Webarchiv Österreich. Retrieved 2021-03-10, from <https://webarchiv.onb.ac.at>. [URL Memento: Wayback Machine]
- Bibliothèque nationale de France. (n.d.). BnF Archives de l'internet [Web page]. BnF; Bibliothèque nationale de France. Retrieved 2021-01-28, from <https://www.bnf.fr/fr/archives-de-linternet>. [URL Memento: Wayback Machine]
- British Library Research Repository. (n.d.). Datasets: UK Web Archive—British Library Dataset Collections. UKWA Open Data/British Library. [https://bl.iro.bl.uk/collections/d09fbc16-7a76-49db-a45f-16a99c30ae3e?utf8=%E2%9C%93&cq=Dataset&sort=score+desc%2C+system\\_create\\_dt+si+desc&per\\_page=100&locale=en](https://bl.iro.bl.uk/collections/d09fbc16-7a76-49db-a45f-16a99c30ae3e?utf8=%E2%9C%93&cq=Dataset&sort=score+desc%2C+system_create_dt+si+desc&per_page=100&locale=en) [BL Research Repository]
- British Library Research Repository. (n.d.). Host Link Graph—JISC UK Web Domain Dataset (1996-2010). UKWA Open Data/British Library. DOI: 10.5259/UKWA.DS.2/HOST.LINKAGE/1
- British Library Research Repository. (n.d.). JISC UK Web Domain Dataset Crawled URL Index. 1996—2013. CDX. UKWA Open Data/British Library. DOI: 10.5259/UKWA.DS.2/CDX/1
- Common Crawl. (n.d.). Common Crawl [Website]. Common Crawl. Retrieved 2021-03-12, from <https://commoncrawl.org>. [URL Memento: Wayback Machine]
- Darcy, E., Charthaigh, C. N., Lalor, M., Sinnott, J., & Shank, C. (Eds.). (2021). In Her Shoes: Stories of the Eighth Amendment (Ireland). Digital Repository of Ireland, Collections. DOI: 10.7486/DRI.wm11nd02p
- Det Kgl. Bibliotek. (n.d.). Netarkivet [Web page]. Det Kgl. Bibliotek. Retrieved 2021-03-09, from <https://www.kb.dk/en/find-materials/collections/netarkivet>. [URL Memento: Wayback Machine]
- International Internet Preservation Consortium. (n.d.). International Internet Preservation Consortium [Website]. International Internet Preservation Consortium. Retrieved 2021-04-22, from <https://netpreserve.org>. [URL Memento: Wayback Machine]

International Internet Preservation Consortium. (n.d.). International Internet Preservation Consortium - Archive-It [Web page]. Archive-It. <https://archive-it.org/home/IIPC>. [URL Memento: Wayback Machine]

Internet Archive. (n.d.). Wayback Machine (Internet Archive) [Web page]. Internet Archive. Retrieved 2021-02-02, from <https://archive.org/web/>. [URL Memento: Wayback Machine]

Koninklijke Bibliotheek. (n.d.). Web archiving [Web page]. Koninklijke Bibliotheek. Retrieved 2022-06-28, from <https://www.kb.nl/en/about-us/expertise/web-archiving>. [URL Memento: Wayback Machine]

Library and Archives Canada. (n.d.). Government of Canada Web Archive [Web page]. Library and Archives Canada. Retrieved 2021-03-08, from <https://www.bac-lac.gc.ca/eng/discover/archives-web-government/Pages/web-archives.aspx>. [URL Memento: Wayback Machine]

Library of Congress. (n.d.). Library of Congress Web Archives [Web page]. Library of Congress. Retrieved 2021-03-26, from <https://www.loc.gov/web-archives/collections>. [URL Memento: Wayback Machine]

Library of Trinity College Dublin. (n.d.). Library of Trinity College Dublin [Web page]. Trinity College Dublin. Retrieved 2022-07-10, from <https://www.tcd.ie/library/> [URL Memento: Wayback Machine]

Littman, J. (2018). Ireland 8th Tweet Ids. Harvard Dataverse. DOI: 10.7910/DVN/PYCLPE

MirrorWeb. (n.d.). MirrorWeb: Your unified compliance platform [Website]. MirrorWeb. Retrieved 2021-11-20, from <https://www.mirrorweb.com/>. [URL Memento: Wayback Machine]

National and University Library in Zagreb. (n.d.). Croatian Web Archive (HAW) [Website]. Croatian Web Archive. Retrieved 2021-10-08, from <https://haw.nsk.hr/en>. [URL Memento: Wayback Machine]

National and University Library of Iceland. (n.d.). Landsbókasafn—The Icelandic web archive. National and University Library of Iceland. Retrieved 2022-03-02, from <https://landsbokasafn.is/index.php?page=icelandic-web-archive>. [URL Memento: Wayback Machine]

National Library of Australia. (n.d.). PANDORA Web Archive [Website]. PANDORA Web Archive, National Library of Australia. Retrieved 2022-01-24, from <http://pandora.nla.gov.au>. [URL Memento: Wayback Machine]

National Library of Ireland. (n.d.). NLI Web Archive—Archive-It [Web page]. Archive-It. Retrieved 2021-03-01, from <https://archive-it.org/home/nli> [URL Memento: Wayback Machine]

National Library of Ireland. (n.d.). NLI Web Archive—Archive-It—Collections [Web page]. Archive-It. Retrieved 2022-07-04, from <https://archive-it.org/home/nli/?show=Collections> [URL Memento: Wayback Machine]

National Library of Ireland. (n.d.). NLI Web Archive—Archive-It—Sites [Web page]. Archive-It. Retrieved 2022-09-22, from <https://archive-it.org/home/nli/?show=Sites> [URL Memento: Wayback Machine]

National Library of Ireland. (2011). NLI Web Archive—General Election 2011. Archive-It. Retrieved 2022-09-27, from <https://archive-it.org/collections/19959>. [URL Memento: Wayback Machine]

National Library of Korea. (n.d.). National Library of Korea Web Resources Archive—OASIS [Web page]. National Library of Korea. Retrieved 2022-09-22, from <https://www.nl.go.kr/oasis/>. [URL Memento: Wayback Machine]

National Library of New Zealand. (n.d.). New Zealand Web Archive [Web page]. National Library of New Zealand. Retrieved 2022-06-01, from <https://natlib.govt.nz/collections/a-z/new-zealand-web-archive>. [URL Memento: Wayback Machine]

Public Record Office of Northern Ireland (PRONI). (n.d.). PRONI Web Archive [Web page]. NI Direct Government Services. Retrieved 2021-02-22, from <https://www.nidirect.gov.uk/services/search-proni-web-archive>. [URL Memento: Wayback Machine]

Public Record Office of Northern Ireland (PRONI). (n.d.). PRONI Web Archive [Web page]. NI Direct Government Services. Retrieved 2021-02-22, from <https://www.nidirect.gov.uk/services/search-proni-web-archive>. [URL Memento: Wayback Machine]

UK Parliamentary Archives. (n.d.). UK Parliament Web Archive [Website]. UK Parliament Web Archive. Retrieved 2021-02-02, from <http://webarchive.parliament.uk/>. [URL Memento: Wayback Machine]

The National Archives. (n.d.). UK Government Web Archive [Web page]. The National Archives (UK). Retrieved 2021-03-17, from <http://www.nationalarchives.gov.uk/webarchive>. [URL Memento: Wayback Machine]

UK Web Archive. (n.d.). SHINE [Web page]. UK Web Archive. Retrieved 2021-11-14, from <https://www.webarchive.org.uk/shine>. [URL Memento: Wayback Machine]; QID: None

UK Web Archive. (n.d.). UK Web Archive [Website]. UK Web Archive. Retrieved 2021-03-18, from <https://www.webarchive.org.uk/ukwa>. [URL Memento: Wayback Machine]

UK Web Archive. (n.d.). UKWA Topics and Themes [Web page]. UK Web Archive. Retrieved 2022-06-19, from <https://www.webarchive.org.uk/en/ukwa/category/> [URL Memento: Wayback Machine]

UKWA Open Data. (n.d.). JISC UK Web Domain Dataset (1996-2013). UK Web Archive/British Library. DOI: 10.5259/UKWA.DS.2/1

University College Dublin. (n.d.). University College Dublin—Archive-It [Web page]. Archive-It. <https://archive-it.org/organizations/1846> [URL Memento: Wayback Machine]

Wikipedia. (2012+). Internet Memory Foundation. Wikipedia (Online/web). Retrieved 2022-06-27, from [https://en.wikipedia.org/wiki/Internet\\_Memory\\_Foundation](https://en.wikipedia.org/wiki/Internet_Memory_Foundation) [URL Memento: Wayback Machine]

Zone-H. (n.d.). Zone-H: Unrestricted information [Website]. Zone-H Unrestricted Information. Retrieved 2021-10-14, from <http://www.zone-h.org/?hz=1>. [URL Memento: Wayback Machine]



## Appendix A: Awareness/Engagement Survey - Recruitment Email Example

### Survey Recruitment Email for Academics

The following contains the text of the email sent to academics in nine universities. The link to the survey is no longer operable. Please note the inclusion of the paragraph in [square brackets] was added after an initial 76 emails were sent.

Dear Professor xx,

I would be most grateful if you would consider participating in this anonymous survey, and sharing this email with students, lecturers, and researchers to which you are associated, for their interest to participate also. This research is being carried out by Sharon Healy, a doctoral candidate at Maynooth University, and is supervised by Prof. Susan Schreibman. This survey is about the awareness of, and engagement with web archives and archived web content in Irish third-level academic institutions.

[A '**web archive**' is a resource that captures and preserves websites, blogs, and web pages, and provides access to view such content, long after it has disappeared from the live web. A web archive differs from a digital archive/ library in so far as a web archive only contains archived websites, blogs, and web pages.]

To participate in the survey, please click here: <https://www.surveymonkey.com/r/WebArchivesIR>

Please note, it is equally important to attain participation from respondents who are **not aware** of, or **do not engage** with web archives, as it is to attain responses from occasional or regular users.

It should take 8-10 minutes to complete this survey. Completion of the survey is voluntary, and participants can withdraw at any time.

The survey is targeted at the following audience:

- Undergraduate students, Postgraduate students
- PhD candidates/students, Postdoctoral associates, researchers or fellows
- Senior Lecturers/Associate Lecturers, Professors/Associate Professors
- Employed researchers in a third-level educational setting or project.

**Purpose of the Study**

For more than two decades, national libraries and cultural heritage organisations have been archiving websites (including blogs), which are then made accessible for current and future research, long after the original website has gone or been changed. However, to date, little is known about the awareness of, or engagement with web archives, and archived web content in Ireland. Therefore, in the context of Irish third-level academic institutions, the aim of this survey is to:

- Investigate the awareness of web archives and archived websites as a resource for study/research
- Generate a better understanding of how and why archived websites are used or not used for study/research

- Explore the challenges and opportunities for using archived websites as a resource for study/research.

Confidentiality

This study is being conducted according to Maynooth University Ethics Committee guidelines and has received their approval. Your confidentiality will be kept at all times. If you have any concerns or would like any further information about this research study, please contact the researcher, [sharon.healy@mu.ie](mailto:sharon.healy@mu.ie) or the supervisor of this research, [susan.schreibman@mu.ie](mailto:susan.schreibman@mu.ie).

Yours sincerely

Sharon Healy

PhD Candidate in Digital Humanities  
GOIPG, Irish Research Council Scholar  
Maynooth University

## Appendix B: Awareness/Engagement Survey - Informed Consent

### **Informed Consent: Awareness of and engagement with web archives, in Irish third-level academic institutions**

The following is the text of the informed consent, which introduced the survey on the SurveyMonkey platform.

#### ***Awareness of and engagement with web archives, in Irish third-level academic institutions***

**Information:** Thank you for taking the time to consider participating in this study. This study is being carried out by Sharon Healy, a doctoral candidate at Maynooth University, and is supervised by Prof. Susan Schreibman. It consists of an anonymous survey and is entirely voluntary. It will take approximately 8-10 minutes to fill out. You may exit at any time during the process of filling out this survey, and your responses will not be recorded. If you wish to participate, simply click Next at the bottom of this page, complete the survey and press submit, and your responses will be recorded as anonymous. If you decide to participate, it is important that you fully understand what is required.

**Purpose of the Study:** For more than two decades, national libraries and cultural heritage organisations have been archiving websites (inclusive of blogs), which are made accessible for current and future research, long after the original website has gone or been changed. However, to date, little is known about the awareness of, or engagement with web archives, and web archived content in Ireland. Therefore, the aim of this survey is to:

- Investigate the awareness of web archives and archived websites as a resource for study/ research
- Generate a better understanding of how and why archived websites are used or not used for study/ research
- Explore the challenges and opportunities for using archived websites as a resource for study/ research.

**What do you have to do?** You must be 18 years of age or over. If you decide to take part, you will be required to complete a questionnaire first on some basic demographic information such as nationality, gender, age, role/ position. Thereafter, you will be required to answer a questionnaire on your awareness or lack of awareness of web archives, and engagement with or lack of engagement with web archives, and web archived content.

**How will the information collected by this survey be used?** This study is being conducted according to Maynooth University Ethics Committee guidelines and has received their approval. Your confidentiality will be kept at all times. All opinions and data will be reported in an aggregated form so that individuals will not be identified. Summaries of the results will be included as part of a PhD dissertation and in other publications associated with this research.

**What if there is a problem?** If you have any concerns or would like any further information about this research study, please contact the researcher, [sharon.healy@mu.ie](mailto:sharon.healy@mu.ie) or the supervisor of this research, [susan.schreibman@mu.ie](mailto:susan.schreibman@mu.ie)

**Who will have access to this data?** This data will not be shared with a third party and will only be processed in a manner compatible with the purposes of this research. Sharon Healy will act as the data controller for all responses gathered and will endeavour to store this data for a period of ten years as outlined in the Maynooth University Research Integrity Policy, after which it will be destroyed. (Please Note: It must be recognized that, in some circumstances, confidentiality of research data and records may be overridden by courts in the event of litigation or in the course of investigation by lawful authority. In such circumstances the University will take all reasonable steps within law to ensure that confidentiality is maintained to the greatest possible extent.)

**Informed Consent:** By clicking Next and submitting this survey, you are also confirming that:

- you are 18 years of age or over
- you have been sufficiently informed about the project
- you are taking part in this research study voluntarily
- you agree to have your responses stored and processed in a manner compatible with the purposes of this research.

**Next**

## Appendix C: Awareness/Engagement Survey - Questions

### Survey Questions: Awareness of and engagement with web archives, in Irish third-level academic institutions

#### About You

These questions allow for the exploration of any trends from the rest of the survey across nationality, age, gender, area of study/research, use of digital research resources.

**Q.1** – What is your nationality?

*Dropdown Box*

Country Index

**Q.2** – Please select your age?

*Dropdown Box*

18-24      25-34      35-44      45-54      55-64      65+      Prefer not to say

**Q.3** – What gender do you identify with?

*Multiple-Choice Box*

Male      Female      Other      Prefer not to say

**Q.4** – Which of the following best describes your current student/ academic/ research position in a third-level academic institution?

*Check Box*

- Undergraduate student
- Postgraduate student
- PhD candidate/student
- Postdoctoral associate, researcher or fellow
- Employed researcher in a third-level educational setting or project
- Senior Lecturer or Associate Lecturer
- Professor or Associate Professor
- Other (please describe)

**Q.5** – Which of the following academic disciplines best describes your primary area of study/ research?

*Multiple-Choice Box*

- Architecture
- Arts (visual, performance, music)
- Business, Economics, Finance
- Computer Science
- Digital Arts, Digital Humanities, Digital Cultural Heritage
- Educational Science

- Engineering Science
- Geography (cartography, hydrology, meteorology, environment)
- Government / Public Administration
- Heritage and Archival Studies
- Humanities (history, archaeology, languages, literature, philosophy, theology)
- Internet Studies
- Law (criminal, civil, common, statute)
- Library and Information Sciences
- Mathematics
- Media/Communications
- Natural Sciences (biology, chemistry, physics, earth sciences, space sciences)
- Political Science
- Social Sciences (anthropology, human geography, linguistics, sociology, psychology)
- Sport and Leisure
- Other (please specify)

**Q.6** – From the list below, please indicate the frequency to which you access/ use the following online/ digital resources to assist with your studies/ research?

	I ALWAYS access/use this resource for my studies/research	I SOMETIMES access/use this resource for my studies/research	I RARELY access/use this resource for my studies/research	I NEVER access/use this resource for my studies/research
World Wide Web	X	X	X	X
Digital Archives	X	X	X	X
Digital Libraries	X	X	X	X
Virtual Research Environments	X	X	X	X

#### Student /Academic/ Researcher Awareness of Web Archives

For more than two decades, national libraries and cultural heritage organisations have been archiving websites (inclusive of blogs), which are made accessible for current and future research, long after the original website has gone or been changed.

Web archiving entails the processes of selecting, capturing, storing, and preserving websites and web pages, and subsequently, ensuring the provision of access to such content for future research and analysis. A web archive then is a resource that stores and preserves captured websites and web content, as well as an access point to view and reference such content.

There are two types of web archives for access:

1. An online public web archive whereby access is available to the general public via the web/internet from any location.
2. A dark web archive which is only accessible onsite in a designated reading room or Library via an onsite portal.

**Q.7** – Prior to commencing this survey, were you aware that the National Library of Ireland archives websites and blogs which are made accessible through the NLI Web Archive – an online public web archive? (<https://archive-it.org/home/nli>)

Check Box

Yes: I was aware / No: I was not aware

**Q.8** – Prior to commencing this survey, were you aware that the National Library of Ireland archived the Irish domain (.ie) in 2007 and 2017 and will soon make it available as a dark archive – only accessible onsite in a designated reading room at the National Library of Ireland?

Check Box

Yes: I was aware / No: I was not aware

**Q.9** – From the list of online public web archive resources below, please indicate your awareness of their existence prior to commencing this survey

Multiple-Choice Box	Yes: I was aware	No: I was not aware
Internet Archive, Wayback Machine <a href="http://archive.org/web/">http://archive.org/web/</a>	x	x
PRONI Web Archive (Public Record Office of Northern Ireland) <a href="https://www.nidirect.gov.uk/services/search-proni-web-archive">https://www.nidirect.gov.uk/services/search-proni-web-archive</a>	x	x
UK Web Archive (British Library) <a href="https://www.webarchive.org.uk/ukwa/">https://www.webarchive.org.uk/ukwa/</a>	x	x
UK Government Web Archive (UK National Archives) <a href="http://www.nationalarchives.gov.uk/webarchive/">http://www.nationalarchives.gov.uk/webarchive/</a>	x	x
UK Parliament Web Archive <a href="http://webarchive.parliament.uk/">http://webarchive.parliament.uk/</a>	x	x
US Library of Congress Web Archive <a href="https://www.loc.gov/websites/collections/">https://www.loc.gov/websites/collections/</a>	x	x

**Q.10** – Are there any other web archives that you are aware of, that are not listed above?

Multiple-Choice Box

Yes / No

**Q.11** – If you answered YES to question Q11 above, would you please enter the names of any other web archive(s) you are aware of (please use commas to separate multiple entries)

Text Box

**Q.12** – To the best of your knowledge, have you ever accessed or used an online public web archive for your personal interest?

Check Box

Yes / No / Unsure

**Q.13** – To the best of your knowledge, have you ever accessed or used an online public web archive, or dark web archive to assist with your studies or research?

Check Box

Yes / No

- If answer is 'NO' to Q.14 above - Go to non-user respondents
- If answer is 'YES' to Q.14 above - Go to user respondents directs to

#### Non-User Respondents

You previously indicated that you have you have NOT accessed or used an online public web archive, or dark web archive to assist with your studies/ research.

This section now looks at some reasons why you do not use a web archive, and whether you might access or use a web archive in the future.

**Q.14** – What are your main reasons to date for not using a web archive for your studies/ research (please tick all that apply)

- I was not aware of the availability of web archives as resources for my studies/ research
- I do not know how to use a web archive for my studies/ research
- I feel that I do not have the technical skills to use a web archive for my studies/ research
- I do not know how to find archived websites relevant to my studies/ research in a web archive
- I do not know how to cite/reference an archived website from a web archive to include in my studies/research
- I am unsure of the credibility or authority of using archived websites as a primary source my studies/research
- I am unsure about copyright implications for using archived web content for my studies/research
- Other reason(s) for not using a web archive for your studies/research (please specify)

**Q.15** – What is the likelihood that you will access or use the NLI Web Archive in the future for your studies/research? (National Library of Ireland online public web archive - <https://archive-it.org/home/nli>)

#### Multiple-Choice Box

Definitely Likely	Fairly Likely	Not Very Likely	Definitively Not Likely	Unsure
-------------------	---------------	-----------------	-------------------------	--------

**Q.16** – From the list of online web archive resources below, what is the likelihood that you will access or use them in the future for your studies/research?

	Definitely Likely	Fairly Likely	Not Very Likely	Definitely Not Likely	Unsure
Internet Archive, Wayback Machine	X	X	X	X	X
PRONI Web Archive (Public Record Office of Northern Ireland)	X	X	X	X	X
UK Web Archive (British Library)	X	X	X	X	X
UK Government Web Archive (UK National Archives)	X	X	X	X	X
UK Parliament Web Archive	X	X	X	X	X
US Library of Congress Web Archive	X	X	X	X	X



**Q.17** – What is the likelihood that you will ever access or use a dark web archive in the future for your studies/research? (only accessible onsite in a reading room or a library via an onsite portal)

*Multiple-Choice Box*

Definitely Likely	Fairly Likely	Not Very Likely	Definitively Not Likely	Unsure
----------------------	------------------	--------------------	----------------------------	--------

User Respondent

You previously indicated that you have accessed or used an online public web archive, or dark web archive to assist with your studies/ research. This section now looks at your engagement with web archives.

**Q.18** – In the context of using a web archive in general, have you accessed or used a web archive for any of the following reasons (please tick all that apply)

- Personal interests
- Historical interests
- Evidential interests
- Research interests
- Cultural interests
- Technical interests
- Design/artistic interests
- Other (please specify)

**Q.19** – In the context of using archived web content (archived websites, blogs, web pages), have you used archived web content from a web archive, for any of the following reasons (please tick all that apply)

- I have used archived web content as a primary source in an academic essay/assignment for my course
- I have used archived web content to document the history of an organisation in an academic essay/assignment for my course
- I have used archived web content as a primary source in a professional research report
- I have used archived web content as a primary source in a professional publication
- I have used archived web content to document the history of an organisation in a professional report/publication
- I have used archived web content as part of my teaching materials for undergraduate students
- I have used archived web content as part of my teaching materials for postgraduate students
- I have used large volumes of archived web content for content analysis/textual analysis/discourse analysis
- I have used large volumes of archived web content for data mining/ topic modelling/ data visualisation
- I have used large volumes of archived web content for network analysis/ geo-spatial analysis
- I have used archived web content for other reasons not listed above - please specify

**Q.20** – Have you ever accessed or used the NLI Web Archive for your studies/ research? (National Library of Ireland online public web archive - <https://archive-it.org/home/nli>)

*Multiple-Choice Box*

Yes / No

**Q.21** – Have you ever accessed or used the following online public web archives for your studies/research?

*Multiple-Choice Box*

	Yes	No
Internet Archive, Wayback Machine	X	X
PRONI Web Archive (Public Record Office of Northern Ireland)	X	X
UK Web Archive (British Library)	X	X
UK Government Web Archive (UK National Archives)	X	X
UK Parliament Web Archive	X	X
US Library of Congress Web Archive	X	X

**Q.22** – Are there any other online web archives that you access or use for your studies/research that is not listed above?

*Multiple-Choice Box*

Yes / No

**Q.23** – If you answered Yes to Q23 above, would you please write down the name(s) of any other online web archive(s) you have accessed or used for your studies/research (please use commas to separate multiple entries)

*Text Box*

**Q.24** - Have you ever accessed or used a dark web archive for your studies/research? (a dark web archive is only accessible onsite in a designated reading room or Library via an onsite portal)

*Multiple-Choice Box*

Yes / No

**Q.25** - If you answered Yes to Q25 above, would you please write down the name(s) of any dark web archive(s) you have accessed or used for your studies/ research (please use commas to separate multiple entries)

*Text Box*

**Q.26** - In your opinion, what is the likelihood that you will access or use a dark web archive in the future for your studies/research?

*Multiple-Choice Box*

Definitely Likely	Fairly Likely	Not Very Likely	Definitively Not Likely	Unsure
----------------------	------------------	--------------------	----------------------------	--------

## Final section

**Q.27** – In your opinion how important is it to archive websites and blogs for current and future research, based on the following values?

### Multiple-Choice Box

	Very important	Fairly important	Slightly Important	Not important	No opinion
Historical value	X	X	X	X	X
Evidential value	X	X	X	X	X
Research value	X	X	X	X	X
Cultural value	X	X	X	X	X
Technical value	X	X	X	X	X
Design/artistic value	X	X	X	X	X

**Q.28** – In your opinion, how important is it to archive websites and blogs based on the following topics?

### Multiple-Choice Box

	Very important	Fairly important	Slightly important	Not important	No Opinion
<b>Indirect Government</b> (websites of agencies deployed by the Irish Government to undertake a task, e.g. Irish Water, Nama)	X	X	X	X	X
<b>Politics</b> (websites/blogs of political parties, political commentators)	X	X	X	X	X
<b>Community Groups/Activists</b> (websites/blogs of clubs, societies, advocacy groups, human rights groups)	X	X	X	X	X
<b>Events</b> (websites/blogs for natural disasters, sporting events, commemoration events)	X	X	X	X	X
<b>Election Campaigns</b> (websites/blogs of candidates, election judicators, commentators)	X	X	X	X	X
<b>Referendum Campaigns</b> (websites/blogs of interest groups, referendum judicators, commentators)		X	X	X	X
<b>Environment</b> (websites/blogs which report on climate change, pollution, conservation)	X	X	X	X	X
<b>Science</b> (websites/blogs which report on advances in medicine, chemistry, physics)	X	X	X	X	X

**Q.29** – In your opinion, do you think web archives will become important as a resource for current, medium or long-term future research in your field?

*Multiple-Choice Box*

	Yes	No	Maybe
<b>Current research</b> (next 5 years)	X	X	X
<b>Medium-term research</b> (5-15 years)	X	X	X
<b>Long-term research</b> (15+ years)	X	X	X

**Q.30** - OPTIONAL: In your opinion, what will be the main challenges to use archived web content for studies/research in your field in the future?

Text Box

**SUBMIT SURVEY**

‘Your Response has been submitted’ Landing Page

Thank You for participating in this research, by filling out this survey. Please feel free to forward the link to this survey to colleagues in other Irish academic institutions.

[www.surveymonkey.com](http://www.surveymonkey.com). The results from this survey are anonymised. However, if you would like to be contacted at some stage in the future for focus groups on using web archives and archived web content please email the researcher, [sharon.healy@mu.ie](mailto:sharon.healy@mu.ie) with your name and position. Please note that providing this information does not compromise the confidentiality and anonymity of the survey. It is impossible to link an email sent to this address to a survey response.

If you would like further information about this research or if you have concerns/questions you would like to discuss about the research, please contact the researcher, [sharon.healy@mu.ie](mailto:sharon.healy@mu.ie) or the supervisor of this research, [susan.schreibman@mu.ie](mailto:susan.schreibman@mu.ie), Maynooth University, Co. Kildare, Ireland

*Please Note: If during your participation in this study you feel the information and guidelines that you were given have been neglected or disregarded in any way, or if you are unhappy about the process, please contact the Secretary of the Maynooth University Ethics Committee at [research.ethics@mu.ie](mailto:research.ethics@mu.ie) or +353 (0)1 708 6019. Please be assured that your concerns will be dealt with in a sensitive manner.*

## Appendix D: Awareness/Engagement Survey - Use of online NLI web archives for studies or research

Table D.1: Breakdown for position and discipline categories of respondents who indicated that they use the online public NLI Web Archive for their studies/research (=23)

Uses of the public NLI Web Archive (=23)	Undergrad	Postgrad	PhD student	Postdoc. researcher or fellow	Employed researcher	Senior Lecturer or Assoc.	Professor or Assoc.	Totals
Architecture		=1						(=1)
Educational Science					=1	=1		(=2)
Geography	=1						=1	(=2)
Humanities	=1		=2	=1	=2	=2	=2	(=10)
Law	=1	=1						(=2)
Media/Communications			=1				=1	(=2)
Natural Sciences							=1	(=1)
Nursing, Midwifery	=1							(=1)
Social Sciences			=1				=1	(=2)
Total	(=4)	(=2)	(=4)	(=1)	(=3)	(=3)	(=6)	(=23)

## Appendix E: Awareness/Engagement Survey - Disciplines for respondents who indicated 'Yes' on the importance of web archives

Table E.1: Discipline categories for respondents (N=239) who indicated 'Yes' on the importance of web archives for current, medium, or long-term future

Discipline Categories	'YES' Current Research	'YES' Medium-Term Research	'YES' Long-Term Research	Total number of respondents per discipline category
Architecture	=1	=2	=2	n=2
Arts	=1	=3	=3	n=3
Business, Economics, Finance	=2	=3	=3	n=6
Built Environment	=0	=0	=0	n=1
Computer Science	=2	=5	=6	n=11
Construction Management	=0	=0	=0	n=1
Dental Science	=1	=1	=1	n=1
Digital Arts/Humanities/Heritage	=2	=3	=3	n=4
Educational Science	=5	=9	=9	n=13
Engineering Science	=6	=11	=13	n=24
Geography	=1	=3	=3	n=4
Government/Public Administration	=1	=0	=0	n=1
Health Studies/Sciences	=5	=6	=7	n=11
Heritage Studies, Archival Studies	=0	=1	=1	n=1
Humanities	=27	=35	=37	n=50
Law	=8	=9	=10	n=18
Mathematics	=1	=1	=1	n=4
Medicine, Biomedical Engineering	=1	=1	=1	n=2
Media/Communications	=2	=4	=4	n=4
Natural Sciences	=5	=10	=12	n=29
Nursing, Midwifery	=6	=6	=5	n=9
Political Science	=3	=5	=5	n=6
Psychotherapy	=0	=0	=0	n=1
Social Sciences	=18	=24	=22	n=33
Totals	(=98)	(=142)	(=148)	(N=239)

**WARCnet Special Reports** is a series of reports related to the activities of the WARCnet network. To ensure the relevance of the publications, WARCnet strives to publish with a rapid turnover. WARCnet Special Reports are edited by Niels Brügger, Jane Winters, Valérie Schafer, Kees Teszelszky, Peter Webster and Michael Kurzmeier. In cases where a WARCnet Special Report has gone through a process of single blind review, this is mentioned in the individual publication.

The aim of the WARCnet network is to promote high-quality national and transnational research that will help us to understand the history of (trans)national web domains and of transnational events on the web, drawing on the increasingly important digital cultural heritage held in national web archives. The network activities run in 2020-23, hosted by the School of Communication and Culture at Aarhus University, and are funded by the Independent Research Fund Denmark | Humanities (grant no 9055-00005B).