

Web Archiving & Preservation Task Force

National Records of Scotland, Edinburgh

Meeting 002

08 November 2018

In Attendance

- Sara Day Thomson, DPC
- Sarah Middleton, DPC
- Leona Taylor, PRONI
- Sean Rippington, University of St Andrews Library
- Maria Praetzellis, Internet Archive
- Carl Davies, BBC Archive
- Els Breedstraet, EU Pub Office
- Helena Byrne, BL
- Carlos Rarugal, BL
- Carly Morel, University of Manchester
- Natalia Klaaser, Lancaster University
- Garth Stewart, NRS
- Sharon McMeekin, DPC
- Susan Corrigall, NRS
- Jenny Hunt, NRS
- Hania Smerecka, Lloyd's Banking Group
- Tom Wilson, UNHCR
- Grainne Loughran, PRONI
- Bechaela Walker, Scottish Council on Archives
- Paul Cunnea, NLS
- Adam Harwood, University of Sussex (via virtual conferencing)
- Christopher Cassells, Scottish Council on Archives
- Patricia Falcao, Tate
- Christopher King, Tate
- Tim Gollins, NRS
- Hugh Hagan, NRS

- Neil Miller, NRS
- Laura Peart, University of Sheffield
- Hrafn Malmquist, University of Edinburgh
- Eilidh MacGlone, NLS

Minutes

09.30 Welcome, tea & coffee

10.00 Meeting Opens & Introduction by Chair, Sara Day Thomson (DPC)

10.10 Welcome by Garth Stewart, NRS

- An overview of New Register House
- National Records of Scotland use their web continuity service to archive websites for the public record, enabling NRS to meet its strategic and statutory collecting obligations. It has archived 143 sites in 18 months
- Born-digital collecting has helped to integrate and change working practices in archival teams at NRS
- NRS partners with NLS and other Scottish government colleagues
- In June 2018, the Internet Memory Foundation was going out of business; an alternative service provider was needed to continue web archiving. Web archiving was able to continue as normal within four months with Archive-It from Internet Archive

10.20 Review of Actions from Previous Meeting

Points of interest from previous meeting and topics that discussed further in today's meeting.

- What is the role of the preservation task force
 - Looking for opportunities where the taskforce can advocate web archiving
- Reaching out to similar funding bodies
 - Trying to embed web archiving requirements and an understanding of web archiving into funding for different projects
- The role of the DPC for this type of support and advocacy
 - Sarah Middleton can be approached for strategies when pushing for advocacy as a group
- Complications with web archive quality assurance
 - The processes and how to integrate quality assurance requirements when inviting third party service providers to tender
- Social media and complex web-based content
 - How to deal with complicated media
- Training and development opportunities
- The preservation and archiving of datasets

10.30 Open Floor to Agenda Items

Anything in particular that attendees would also like to discuss?

- How to attract more users to web archives
- How to determine the audience and users of your web archives

10.45 Open Discussion Session 1

Attendees break into four groups, allowing for brainstorming and open discussion (Chatham House rules) around the following ideas:

- Advocacy and monitoring for legislation opportunities
- Quality assurance and requirements for procurement
- Social media and other complex forms of web media
- Attracting users and user awareness

12.10 Discussion and summary of group work

• **Advocacy/Legislation**

- Talking to local authority groups about web archiving; what it is and what it isn't
- Determining the infrastructure in place in public bodies and how web archiving fits in
- The need for continued education/guidance/training; even after the idea of web archives has been introduced and is already in place. Not just a one-off event, people leave/change roles so it needs regular reviews and refreshing of web archiving practices within the organisation
- Question: how is web archiving handled within a local authority?
 - Currently revisiting the procedures in place to make sure obligations are met.
 - Checking current education/guidance to see if the local authority is moving in the right direction and ensuring that web archiving becomes a part of that.
 - The disconnect with trying to archive web sites within the current archiving environment, you need to be able to archive different materials within their correct archive.
 - How to make web archiving an intrinsic part of the current archiving process. Web archiving is a growing and important area.
 - People have the perception that the web is a non-archive, many misunderstandings.
 - The need to develop a proper message to push out to popular forums just what web archiving is/means and what the obligations are under current legislations.
 - Building web archiving into the process of current system of archiving physical stuff.
- Highlight extra resources that DPC can help with i.e. executive guide to digital preservation, extracting key messages as to why web archiving is important
- DPC can help point to other useful resources such as guidebooks and handbooks, within the DPC knowledge base.
 - Concise and digestible descriptions of the many aspects of digital preservation
 - There have been successes because they have been able to point people to good guidance, good help, good advice, to help them make the right decisions. Also, because they can meet people face to face and create a network, a forum for sharing experiences both good and bad. You need to do more than read a guide, you need more engagement
- The LDL review to persuade government to relax restrictions on how users access legal deposit web archives. The open access content is miniscule in comparison to what has been archived.
 - How do we provide advocacy to government to relax these restrictions?
 - How do we collectively, or individually, influence government regarding these issues?
 - Could write to MPs using a standard letter.
- DPC could write on behalf of its membership

- **Attracting users and user awareness**

- Attracting users, how to contact users and let them know what useful things they can see
- Once you have them interested, how do you educate them in actually using web archives?
 - Teaching them how to access and how to process the information.
- Make them aware of any legal/rights issues
- How to structure web archives to make it more usable and accessible
- How to gather accurate user stats to get a better idea of what types of users you have
- Digital literacy is in decline because there is decline in access to digital material
 - Educate on how general systems and I.T infrastructure work
- Students tend to misunderstand what software is as opposed to apps
- Attract users by trying to engage them on a particular subject and how a web archiving can be used to research that subject
- Making sure you're offering training or learning resources
- Use-cases are a good way of reducing barriers to accessing archived websites, users will understand the workflows of other users and apply them to their needs
- Identifying particular audiences, who you would like to target e.g. genealogists
- Identifying frequent/ infrequent events where users may see the benefits of web archiving
- Building web archiving into existing outreach programs e.g. digital tourism or digital literacy programs, webinars, blogs
- Produce web-archiving guides, what it is, how it is accessed stored etc.
- Glossaries, simple things like a breakdown of a URL. Generally demystifying web structures and the processes that enable WA
 - Gamification of the web archiving process
 - Collaborations and promotion of WA
 - Practical resources, worksheets for students
 - Resources for helping people to put on web archiving workshops
- Could a follow on task force sub group be created to determine what materials could be easily shared (within the taskforce or publicly)
- Having good examples would be enlightening
- Having PHDs/researchers interested in the web archive collection and then figure out from their experiences what content is/isn't easily accessible and then see how services could be improved.

- **Quality Assurance and requirements for procurement**

- Procurement and QA are intermingled in a sense because if you're looking for a third-party service, you want know that they are providing a quality service
 - How do you do QA on 1000s or even millions of webpages
- How does QA come into the procurement process in terms of contract specifics?
- What requirements people have or look for and whether DPC members can share these requirements
- What are people currently doing for QA
 - Could we see examples of people's workflows and how they tackle QA issues
- Exit strategies
 - If you are using third parties, how do you transition from one provider to another and what are the pitfalls.
 - If WARC files are returned, what about the extra information or software that is needed to use the WARC files. The metadata, log files, curations tools.
- Avoiding vender lock-in

- If a service provides access to web archives, how do you continue giving users access when that service stops
- Generally, what things to look out for in T+Cs
- Internal legal advice can be somewhat disconnected from the process of procurement
 - Is this the case with other members?
- Active contract management in many Public institutions tends to be weakness as they will not engage with services that do not strictly meet their contract requirements, which means there is little flexibility
 - Not many vendors in web archiving, so there is little space to move from one vendor to another
- How technology is constantly changing, if the ability to capture websites decreases, how do you build these requirements into your contract
 - At tender time, it may be best to select your vendor by looking at their stance in areas of particular interest (e.g. ability to capture emerging formats).
 - Could be difficult to enforce future requirement in T+Cs
- Social media captures, how to do QA on frequently updated archives
- tender documents for public institution procurement should be publicly available

12.45 Lunch: Tour of NRS by Garth Stewart and Susan Corrigan

13.15 Discussion Continues

- **Social media and other complex forms of web media**
 - Distinctions and challenges of email archiving and web archiving, looking at email preservation task force
 - MirrorWeb and the issue of social media data being retained by third-party services
 - Distinction between preservation and access for third-party services
 - The role of long term preservation should be taken up by the organisation
 - Twitter
 - Difference between csv and API capture
 - Could we develop a canonical form of archiving tweets
 - Possibility of developing a labelling system for twitter, labelling individual tweets
 - There is a research group called Documenting the Now, that researches social media. A researcher in the group is currently developing a system that could enable twitter users to label their tweets so that they can be used appropriately by archivists and researchers
 - GDPR and EC guidance
 - Challenges of collecting social media and dynamic web content
 - Could there be a separate session on collecting social media as it is growing in importance
 - There are still lots of grey areas surrounding social media archiving
 - Could be a future theme for a DPC task force
 - Managing expectations for researchers when using social media archives

13.25 Tour of NRS by Garth Stewart and Susan Corrigan, followed by Comfort Break

14.30 Show & Tell of Web Archiving at NRS by Garth Stewart (Web Archivist)

15.00 Open Discussion Session 2

Groups break up and move around to any of the four areas of discussion.

15.30 DPC present beta version of WAPTF Knowledge Exchange Platform on DPC Website

- DPC presents a dedicated part of their web platform for sharing resources and links that can be accessed by this taskforce, a place to continue the conversation about web archiving showcased by Sarah Middleton
- Lots of web archiving areas need standardising, DPC can help bring us together and tackle it collectively as opposed to working on it individually and disjointedly.
- It's a working area, members can add what they like
 - can it be made clear that it's member's added content as opposed to official DPC content
- Area preserves Chatham House rules

15.45 Set Date and Venue for Next Meeting

Start scoping actions for following meeting, open tables

- Drafting letters to MP, scoping points that we might want to address in the letter
- Scoping what resources could be shared for outreach
- Social media group, discuss a standardised way or method for containing social media, technical paths
- Procurement process with perhaps examples from members

16.00 Close of session

Minutes taken by Carlos Rarugal with thanks from the DPC and the rest of the WAPTF

Published 13/05/2019