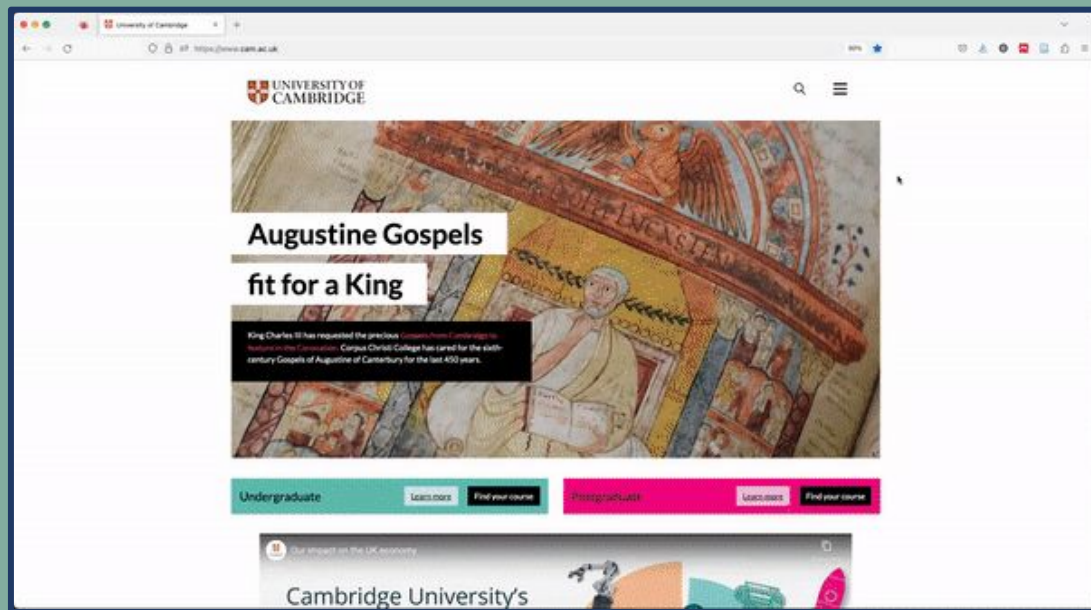


# Back to Class

## Capturing the University of Cambridge Domain

Caylin Smith | Head of Digital Preservation

Leontien Talboom | Technical Analyst





UNIVERSITY OF  
CAMBRIDGE

Welcome to the University of Cambridge [World Wide Web server](#), run by the [Computing Service](#). Please send comments and corrections to [webmaster@ucs.cam.ac.uk](mailto:webmaster@ucs.cam.ac.uk) or use the [comment form](#).

A [WWW cache server](#) is available on a pilot trial basis. [Access is limited to use by client systems within the University of Cambridge (xxx.cam.ac.uk hostname).]

## Finding information

- [Search documents on this WWW server](#).
- [Search WWW servers Internet-wide](#).
- [Postal and email addresses and telephone/FAX numbers](#) for people and all Departments, Colleges, and other institutions within the University of Cambridge.
- [Department and College information servers in the University of Cambridge](#).
- [Recent changes and additions on this server](#).

## The University of Cambridge

- [About the University](#). General information, publications and research interests.
- [Information for prospective students](#), with undergraduate and postgraduate prospectuses, and also [Continuing Education](#), including [International Summer Schools](#), Public and Legal Programmes, and the [Cambridge Programme for Industry](#).
- [Information for students and staff](#).
- [University of Cambridge Local Examinations Syndicate](#)
- [Finding people](#) in the University of Cambridge, including email address information.
- The [event calendar](#), and Departmental and Faculty [seminars](#).
- University-wide [announcements](#), [advertisements](#), and [events newsgroups](#).
- [University Societies](#) pages.
- Information about and provided by the [Computing Service](#).
- Information about [libraries in the University of Cambridge](#) (including telnet access to the [Cambridge University Library catalogue](#)).

## The Cambridge area

- Information about [the Cambridge area](#).
- [Cambridge <-> London train timetables](#).
- [On-line information services in the Cambridge area](#).

## Other information resources

- [Department and College](#) information servers in the University of Cambridge.
- [Academic Information Resources](#) - access to worldwide resources.
- Servers in the [UK and the rest of the world](#)
- The [Computer Laboratory WWW server](#), including the world-famous [networked coffee machine](#).

University of Cambridge, Cambridge, UK  
Telephone +44 1223 337733  
[webmaster@ucs.cam.ac.uk](mailto:webmaster@ucs.cam.ac.uk)

# The Trojan Room Coffee Machine

One of the more unusual outcomes of the work in the Computer Laboratory on multimedia systems was the world's first 'webcam'. A lashed-up camera watched a coffee pot shared by a group of people and placed a live image displaying the status of the coffee pot on the desktops of all group members. It alerted people at some distance when a freshly brewed pot was available and unnecessary journeys were thus avoided. **In 1993 it was obvious that web browsers could be used to display the image and the camera was connected to the Internet. The picture of the coffee pot became visible to Internet users and gained in popularity at an amazing rate.**

*Source: Cambridge Computing, The First 75 Years*

## The Trojan Room Coffee Machine



[Click here](#) for an up to date picture of the Trojan Room coffee machine.

### About this service

This service was created as the first application to use a new RPC mechanism designed in the Computer Laboratory - MSRPC2. It runs over MSNL (Multi-Service Network Layer) - a network layer protocol designed for ATM networks.

A video capture board in an Acorn Archimedes grabs one frame every second and the WWW server requests a frame from it using MSRPC2 (indirectly). Each frame is JPEG encoded by the Archimedes.

The online coffee pot actually predates our World Wide Web server by some time. If you want to know how it all came about, take a look at its (non-technical) [biography](#).

The University of Cambridge Computer Laboratory home page is [here](#).

### Note:

- The lights in the Trojan Room aren't always switched on, but we try to leave a small lamp pointing at the coffee pot so you can see it at night.
- Apart from the fares to get here, you need to be a member of the coffee club before partaking.

[Daniel.Gordon@cl.cam.ac.uk](mailto:Daniel.Gordon@cl.cam.ac.uk)

### Local radio report

On 11th November 1994, we were visited by a reporter from our local radio station, BBC Radio Cambridgeshire, to make a report on this service. Naturally we connected a radio to one of our workstations and relayed the broadcast over our local network. The transmission was also recorded digitally, and now [you can hear it too](#).

We are grateful to BBC Radio Cambridgeshire for giving permission to put this audio file on the Web.

Martyn.Johnson@cl.cam.ac.uk

# How did CUL get here?

| 2003                                                                                                                                                                                                                                                                                                                                                                                   | 2013                                                                                                                                                                                                | 2020                                                                                                                                                                                                                                                                                                                                                                | 2021                                                                                                                                                                                                                                                                                                                                   | 2022                                                                                                                                                                                                                                                                                                                     |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p><b>Legal Deposit Libraries Act 2003</b></p> <p>Updated UK legal deposit legislation</p> <p>British Library; National Library of Scotland; National Library of Wales; Library of Trinity College, Dublin; Bodleian Libraries, University of Oxford; Cambridge University Library</p> <p>Provisions for the deposit of non-print publications, including content published online</p> | <p><b>The Legal Deposit Libraries (Non-Print Works) Regulations</b></p> <p>Deposit of UK published works in digital formats</p> <p>Legal Deposit Web Archiving Sub Group for the UK Web Archive</p> | <p><b>University Digital Presence Programme</b></p> <p>The production of new design system and pattern library</p> <p>Training and support for improved ways of creating content</p> <p>The development of new ways to get and re-use useful data sources for websites from across the University</p> <p>Better governance and support for user-centered design</p> | <p><b>CUL Digital Preservation Programme</b></p> <p>Central place to store and manage digital materials</p> <p>Improve end-to-end workflows and reduce manual steps</p> <p>Migrate to cloud, optimised where possible, reducing operational costs</p> <p>Improve search, discovery, and access across digital collection materials</p> | <p><b>Archive of Tomorrow Project</b></p> <p>Archiving health-related information published online</p> <p>Researching ethics of libraries providing access to false, misleading, malicious, etc. information</p> <p>Researching user needs for web archive content</p> <p>Bonus! Re-start of CUL Web Archiving Group</p> |

# UKWA Content Scope

## UK Top-Level Domains

.uk                      .cymru  
.scot                    .wales  
.london                .gb

## UK hosted with Non-UK Domain

Hosted in the UK  
  
Identified with geoip lookup  
  
uses other kind of domain  
(.com, for example)

## Published outside UK

Project University is involved in with international partners  
  
Permission sought from copyright owner

# How did CUL get here?

| 2003                                                                                                                                                                                                                                                                                                                                                                                   | 2013                                                                                                                                                                                                | 2020                                                                                                                                                                                                                                                                                                                                                                | 2021                                                                                                                                                                                                                                                                                                                                   | 2022                                                                                                                                                                                                                                                                                                                     |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p><b>Legal Deposit Libraries Act 2003</b></p> <p>Updated UK legal deposit legislation</p> <p>British Library; National Library of Scotland; National Library of Wales; Library of Trinity College, Dublin; Bodleian Libraries, University of Oxford; Cambridge University Library</p> <p>Provisions for the deposit of non-print publications, including content published online</p> | <p><b>The Legal Deposit Libraries (Non-Print Works) Regulations</b></p> <p>Deposit of UK published works in digital formats</p> <p>Legal Deposit Web Archiving Sub Group for the UK Web Archive</p> | <p><b>University Digital Presence Programme</b></p> <p>The production of new design system and pattern library</p> <p>Training and support for improved ways of creating content</p> <p>The development of new ways to get and re-use useful data sources for websites from across the University</p> <p>Better governance and support for user-centered design</p> | <p><b>CUL Digital Preservation Programme</b></p> <p>Central place to store and manage digital materials</p> <p>Improve end-to-end workflows and reduce manual steps</p> <p>Migrate to cloud, optimised where possible, reducing operational costs</p> <p>Improve search, discovery, and access across digital collection materials</p> | <p><b>Archive of Tomorrow Project</b></p> <p>Archiving health-related information published online</p> <p>Researching ethics of libraries providing access to false, misleading, malicious, etc. information</p> <p>Researching user needs for web archive content</p> <p>Bonus! Re-start of CUL Web Archiving Group</p> |

# University Archives at the University Library

- Official institutional record of the University
  - Evolving state of what are considered official record(s) as well as formats used
- Held at Cambridge University Library
- Guided by Records Retention Schedule and Collection Policy
- Content related to:
  - Governance
  - Teaching, learning, and research
  - Asset management
  - Student administration
  - External relations management

The University Archives is responsible for the selection, transfer and preservation of the internal, administrative records of the University of Cambridge, dating from 1266 to the present, and for making them available for administrative and research purposes. **The University Archives aims to provide a full and richly varied picture over time of the University's organisation and governance, its key functions and activities, major developments and achievements.**

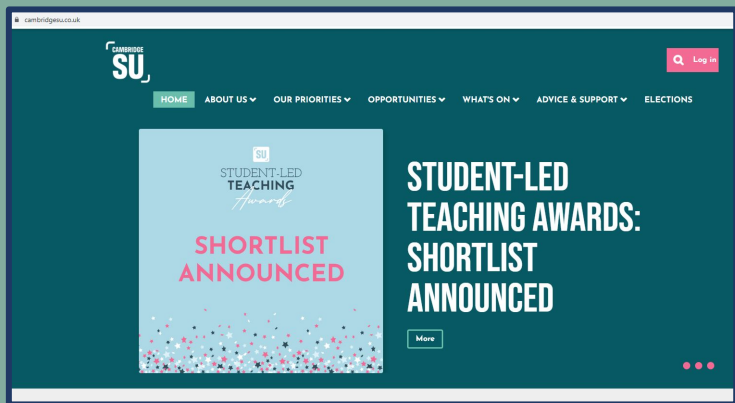
Source: University Archives webpage

## Some stats on cam.ac.uk

3,200 unique websites

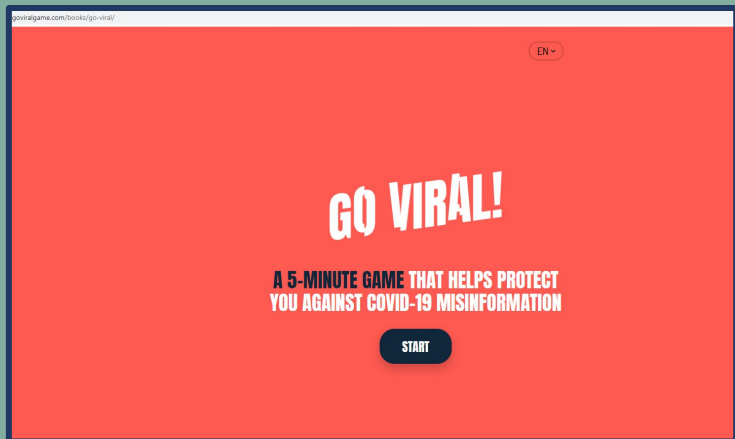
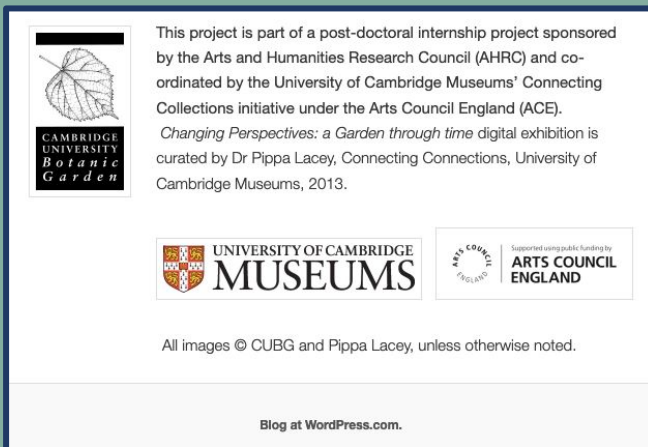
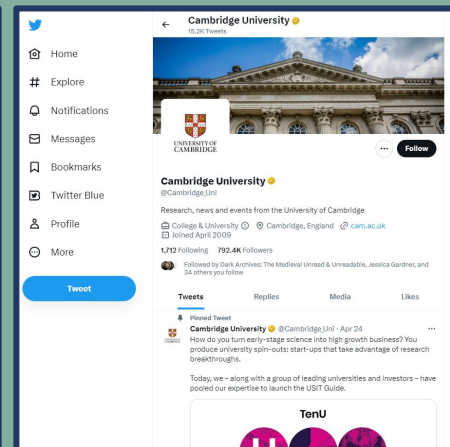
Over 2 million web pages

Over 1,300 web editors



# Wait! There's (likely) more!

- Research project websites and pages
- University content on .com websites
- College websites
- Personal websites of people with a University connection
- Social media



# To Summarise

- Prevent a black hole of information in the University Archives
- Increase in risk of content loss
  - Website redesign
  - General obsolescence
- Increase in dynamic websites
- Improve records management practice amongst staff
- UKWA
  - Some content out of scope
  - Infrastructure
  - Access restrictions

About this site | University of Cambridge  
04/03/2013, 09:48:54

About this site

[Home](#) [Privacy policy](#) [Accessibility](#) [Cookies](#) [Terms and conditions](#) [Connect with us](#)

About this site

[Privacy policy](#)

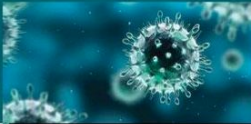
[Accessibility](#)

[Cookies](#)

[Terms and conditions](#)

[Connect with us](#)

About this site

  
  
Spotlight on infectious diseases

"A new infectious disease that goes on to threaten the health of the global population is a likely scenario"

University of Cambridge

  
About the University

  
Study at Cambridge

Latest News

Announcing the greatest of the stars

14 February 2013

The stars and planets have been the subject of our website for many years. We have been working hard to make it more accessible to our visitors. We will be working hard over the coming year and beyond to upgrade as many of our sites as possible.

The homepage and core pages of the University of Cambridge's website (including our news and research pages) are maintained by a group of web teams.

Aside from the core site, the University hosts over 600 different websites which cover Colleges, departments and research groups, and the University Administration Service. Our search engine covers all of these sites.

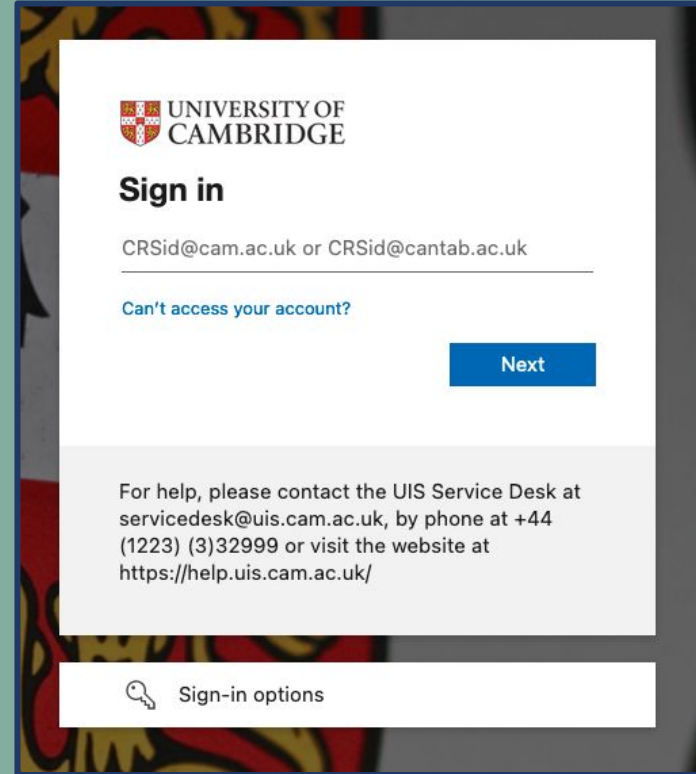
### 2013 redesign

We have recently undergone a redesign of our web templates in order to make our pages more accessible on mobile and tablet devices. This is the culmination of over two years' work and marks a significant change in the way that websites are delivered by the University of Cambridge. We will be working hard over the coming year and beyond to upgrade as many of our sites as possible.

[Send feedback, comments and suggestions](#)

# Difficulties Around Capturing Cambridge Domain

- Sub domains are not picked up by main website crawl
- Understanding the scope
- Many projects and events have their own URL
- Intranet
  - Two-factor authentication
  - Not in scope for legislation
- Data driven and dynamic content
- Social media



The screenshot shows the University of Cambridge Sign in page. At the top is the University of Cambridge crest and name. Below is the 'Sign in' heading. A text input field contains the email addresses 'CRSid@cam.ac.uk or CRSid@cantab.ac.uk'. Below the input field is a blue button labeled 'Next'. To the left of the button is a link that says 'Can't access your account?'. At the bottom of the page is a footer section with a key icon and the text 'Sign-in options'. The background of the page is a patterned design.

UNIVERSITY OF  
CAMBRIDGE

**Sign in**

CRSid@cam.ac.uk or CRSid@cantab.ac.uk

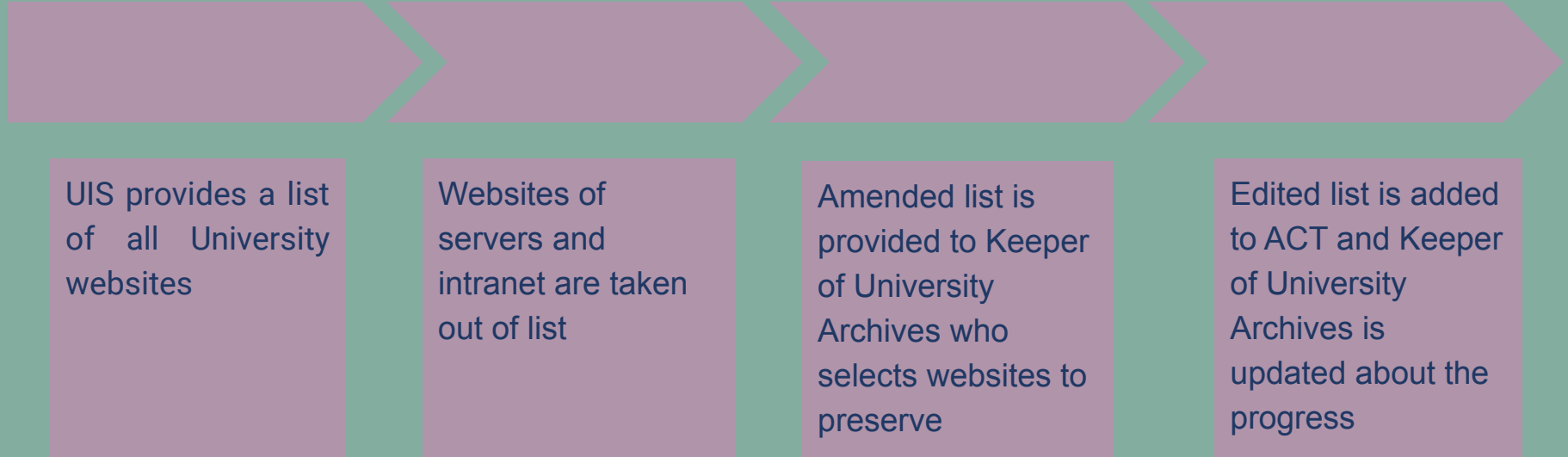
[Can't access your account?](#)

**Next**

For help, please contact the UIS Service Desk at servicedesk@uis.cam.ac.uk, by phone at +44 (1223) (3)32999 or visit the website at <https://help.uis.cam.ac.uk/>

 Sign-in options

# Workflow



# Intranet

- Nearly every department at Cambridge has an Intranet.
- Questions around how to capture it, especially from an ethical point of view.
- Keeper of University Archives is able to help with the selection process
- We were trying to find a tool that was easy to use and open-source
- From a technical point of view: two things that have worked for us: HTTrack and Conifer

Log in

These pages are only accessible to members of the University of Cambridge with a Raven account.

[Log in here](#)

## Cambridge University Libraries Intranet

Home

All Staff Briefing

Departments & Affiliates

HR

Key Documents

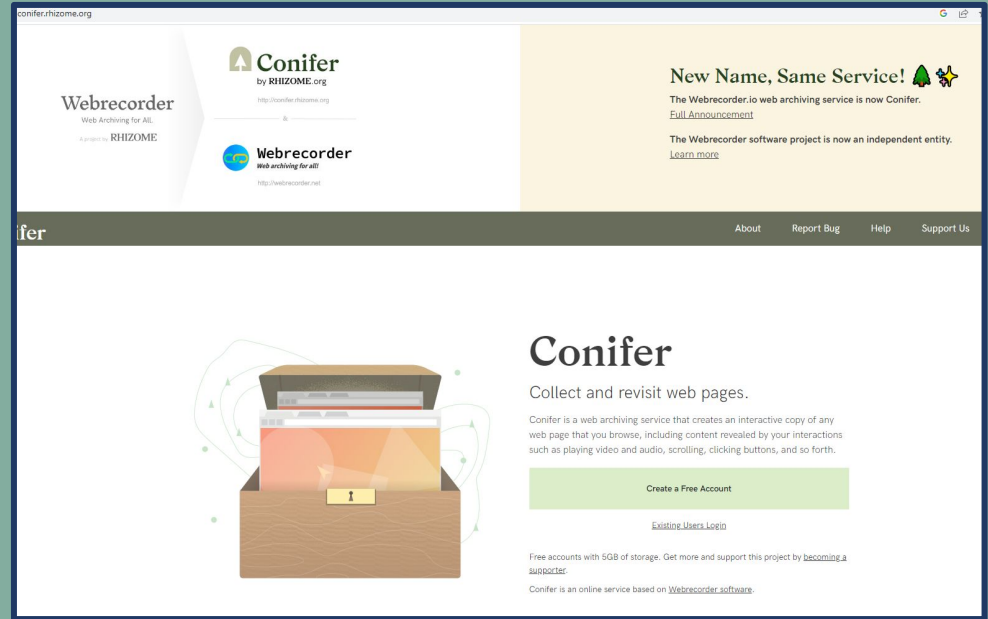
Library communications

Staff Groups



# Conifer

- Web-based tool to capture websites
- Creates WARC files
- No downloads necessary
- Easy interface
- Time-consuming
- Difficult to keep track



# HTTrack

- Website copier
- Requires a download
- Does not create a WARC file
- Easy to browse capture
- Automatic crawl of intranet

In progress:

Transferring data..

Information

|                     |                         |                |                |
|---------------------|-------------------------|----------------|----------------|
| Bytes saved:        | 453,34MiB               | Links scanned: | 285/721 (+437) |
| Time:               | 16min20s                | Files written: | 682            |
| Transfer rate:      | 2,50KiB/s (442,01KiB/s) | Files updated: | 590 (86%)      |
| Active connections: | 1                       | Errors:        | 5              |

☒ Actions:

|          |                               |             |                 |
|----------|-------------------------------|-------------|-----------------|
| scanning | .tps://www.in...and-standards | <div></div> | <div>SKIP</div> |
|          |                               | <div></div> | <div>SKIP</div> |
|          |                               | <div></div> | <div>SKIP</div> |
|          |                               | <div></div> | <div>SKIP</div> |
|          |                               | <div></div> | <div>SKIP</div> |
|          |                               | <div></div> | <div>SKIP</div> |
|          |                               | <div></div> | <div>SKIP</div> |
|          |                               | <div></div> | <div>SKIP</div> |
|          |                               | <div></div> | <div>SKIP</div> |
|          |                               | <div></div> | <div>SKIP</div> |
|          |                               | <div></div> | <div>SKIP</div> |
|          |                               | <div></div> | <div>SKIP</div> |
|          |                               | <div></div> | <div>SKIP</div> |
|          |                               | <div></div> | <div>SKIP</div> |

< Back

Next >

Cancel

Help

## Next Steps and Get in Touch

- Continue to work with archivists to scope web content for University Archives or other areas of collecting (e.g., project websites)
- Assess web archiving tools and services against requirements
- Create a proposal and plan for immediate action
  - Address content that could disappear as part of Cambridge University website redevelopment
- Create a proposal and plan for business as usual web archiving

Caylin Smith | [cs2059@cam.ac.uk](mailto:cs2059@cam.ac.uk)

Leontien Talboom | [lkt39@cam.ac.uk](mailto:lkt39@cam.ac.uk)