

Web Archiving Strategy

Overview

Introduction

Web archiving is the process of creating reliable copies of web-based content for long term preservation. Heritage Collections hold records that document over 400 years in the life of the University, including records of institutions, organisations, and departments that have merged with the University or are no longer operating. Like many other institutions, the University's communications are increasingly disseminated via online channels. As such, web archiving is an integral component of the University's vision for a library and university collection that is at the heart of education, research and engagement at Edinburgh.

Purpose

This document outlines the University of Edinburgh's strategy for archiving selected web content as part of a robust web lifecycle management programme. This document:

- describes the roles and responsibilities of University staff throughout the lifecycle of web content
- outlines the scope of the University's web archiving activity
- identifies the primary means by which web sites are archived for long-term preservation
- describes the principles that guide this work

Roles & Responsibilities

Library & University Collections

- Works with the Website Communications Team to deliver the Web Lifecycle Management Programme on behalf of the University;
- Oversees appraisal with input from Records Management;
- Advises on best practice for creation of web-based content for long-term preservation;
- Develops and publishes guidance and training on web archiving for staff and students;
- Supports the use of the archived web for research and compliance;
- Monitors and supports compliance with the principles outlined here.

Records Management

- Advises on retention schedules and assists with identification of 'golden copy' web content;
- Ensures compliance with legal frameworks and relevant information governance best practice.

Learning, Teaching, Web

- Integrates preservation principles into website creation and development activities;
- Advises on emerging web technologies and their implications for University web archives;
- Advises on planned web lifecycle management activities.

Collection Development

Selection

Heritage Collections actively collects the archives of the University of Edinburgh and related or predecessor institutions, as outlined in the University Archives Collections Development Policy (Appendix G of the [Collections Management Policy 2020-2030](#)). In order to support this function, we seek to preserve the following web content:

- Web content created, owned, or published by University departments, staff, or students in the course of carrying out University-related business, functions, or activities;
- Web content that reflects functions or activities associated with the University's administration, teaching, research, and student life;
- Externally created or owned web content from organisations, groups, or individuals whose archives we are committed to preserve;
- Websites with unique content that documents the work and/or thought of a person or organisation affiliated with the University;
- Websites with content that is not likely to be duplicated in the paper records of an individual or organisation affiliated with the University.

Capture

Content published as part of the University web estate is predominantly archived through the UK Web Archive. By law, all UK print and digital publications - including websites - must be deposited with the British Library and by request to the other five Legal Deposit Libraries (the National Library of Scotland, National Library of Wales, Bodleian Libraries, Cambridge University Libraries and Trinity College, Dublin). The UK Web Archive (UKWA) is a partnership of these six UK Legal Deposit Libraries and aims to collect a copy of all UK websites at least once per year.

Some web content cannot be captured through the UKWA due to legal or technical restrictions. Web content that falls outwith the UKWA geographical remit (non-UK content) and web pages consisting primarily of moving image or sound will not be captured in the UKWA. Content that is only made available to a restricted group of people, such as all the Intranet and Sharepoint pages published by the University, also will not be captured by the UK Web Archive.

When web content in scope for the University Archive cannot be captured in the UK Web Archive, or when the Archive requires a local copy of a particular web page or pages, it is captured using alternative approaches. Capture may be carried out using an open-source tool for manual capture, such as the [WebRecorder tool suite](#), or a third-party vendor with specialist capabilities may be contracted.

Local captures of web content are stored as Web Archive Collection (WARC) and Web Archive Collection Zipped (WACZ) files and ingested into the University's digital archive in line with the University's Digital Preservation Policy.

Information about the tools and services used to capture University content is shared through the Heritage Collections web pages.

Access

University web content archived through the UKWA is made available through a collection page on the UKWA's public interface. Access to websites that have been archived in the UKWA is by default restricted to users at computer terminals onsite in Legal Deposit Libraries. The University has granted a license which allows any archived University pages to be accessed outwith the Legal Deposit Library network. This license automatically applies to all content on the main web domain (<https://www.ed.ac.uk>) and associated subdomains, and is manually applied to University-managed URLs located outside the main domain.

Requests to access local captures of web content should be made through the Heritage Collections Research Service (HeritageCollections@ed.ac.uk), and Research Services staff will make captures available to view.

In certain exceptional circumstances, access to captures may be restricted in line with existing University Takedown Policies. These restricted captures will still appear on the collection page, but access will only be possible at computer terminals onsite in Legal Deposit Libraries or, for local manual captures, onsite in the Centre for Research Collections. In extraordinary circumstances when personal data or other legal breaches have been discovered in the web archive, University archive staff will urgently trigger a review and full suppression of content in breach. When the content in breach sits in the UK Web Archive, procedures for review by the UK Web Archive will be initiated, details of which can be found on the UK Web Archive's website (https://www.webarchive.org.uk/en/ukwa/info/notice_takedown/).