



Borthwick Institute for Archives

Web Archiving Policy



OVERVIEW

1.1 Purpose

- 1.1.1 This policy outlines the Borthwick Institute for Archives' strategy for the archiving of selected websites and their content. These archiving activities occur in collaboration with website owners and the appropriate copyright holders, where applicable.

1.2 Mission

- 1.2.1 Preservation of valuable web content for future generations is integral to the stewardship role of the University of York. As such, the Borthwick manages a service for the preservation of web content that has organisational, informational, historical and/or cultural value. This service is part of the Borthwick's wider mission to support and expand the University of York's cultural endeavour and contribute to human understanding through collecting archives, preserving them and making them widely available for research to all people, now and in the future. This policy aligns and should be read in parallel with the Borthwick's Digital Preservation Policy.

1.3 Principles

- 1.3.1 The Borthwick's approach to web archiving is guided by the following principles:
- We collect materials that align with criteria set out by the Collections Development Policy for the Borthwick Institute for Archives. Similarly, the

preservation of materials approved for web archiving aligns with the principles and requirements outlined in the Borthwick's Digital Preservation Policy.

- Websites are included as part of our web archiving program with the site owner's consent and with written permission from the copyright holder, where applicable.
- The Borthwick's web archiving program follows principles and techniques of non-intrusive harvesting in order to minimise server impact where live sites are hosted. In the event that servers are affected, we provide site owners with clear mechanisms for reporting issues.
- The Borthwick does not archive password-protected content, unless by special arrangement in collaboration with the site owner.
- Up-to-date information about the tools and services used to capture, host, and manage the Borthwick's web archives is made publicly available via the Borthwick's own web pages, alongside mechanisms for getting in touch to raise issues or ask questions.
- We consider web content to be published records and, as a result, provide open access by virtue of the information already having been placed in the public domain.

COLLECTION DEVELOPMENT

2.1 Selection and Capture

2.1.1 The Borthwick actively collects web content created by and about the University of York and web content created externally to the University. In selecting materials for permanent preservation, the Borthwick seeks:

- Websites created, owned, or used by university departments, staff, or students in carrying out university-related business, functions, or activities;
- Websites that reflect functions or activities associated with the university's administration, teaching, research, and student life;
- Externally created or owned websites from organisations, groups, or individuals whose archives we are committed to preserve;
- Websites that fill gaps in our existing collections;
- Websites with unique content that documents the work and/or thought of a person or organisation;
- Websites with content that is not likely to be duplicated in an individual or organisation's paper records.

2.1.2 The identification, appraisal, and selection of university web content is informed by the University Archives Collection Policy. The identification, appraisal, and selection of externally-created web content is informed by the Collections Development Policy for the Borthwick Institute for Archives.

2.1.3 The Borthwick makes every effort to produce exact captures of targeted websites,

but acknowledges that some features of contemporary sites and the technical limitations of tools may result in the incomplete preservation of a given site's content, appearance, and/or functionality.

For more information on the Borthwick's methodology for the acquisition of content, please see <https://www.york.ac.uk/borthwick/holdings/Web-archives/>

2.2 Description and Access

- 2.2.1 By default, archived websites will be made freely available to the public following a minimum six-month embargo period. Access will be provided via the University of York's [Archive-It partner page](#), where website-level metadata is added to allow browsing and full-text search. Where appropriate and in discussion with content owners, selected content may be subject to longer embargo periods. Relevant links will also be included in collection guides and/or catalogue records.

RIGHTS AND OWNERSHIP

3.1 Intellectual Property Rights

- 3.1.1 Materials included as part of the Borthwick's Web Archive are collected to ensure long-term access for research purposes and private study. The Borthwick strives to respect the rights of content owners and to follow professional best practices for intellectual property rights management in website preservation.
- 3.1.2 To address the rights of content owners, the Borthwick and its web archiving service provider have taken the following steps:
- All preserved materials are prominently labelled as an 'archived web page' with information about the capture so as to avoid confusion with 'live' websites.
 - Content owners may request that portions of their site be suppressed from public view and can choose to opt out entirely from captures.
 - When the Borthwick preserves the websites of private individuals, organisations, or associations (including University of York staff and student organisations), it does so with the site owner's consent and with written permission from the copyright holder, where applicable.
 - Information about the Borthwick's web archiving program is made publicly available via our privacy notice and web pages in order to inform possible content owners of our activity and their right to opt out or suppress content.
 - The Borthwick will respect and adhere to robots.txt exclusions, but may contact site owners to request changes to rules or seek permission to override them where such exclusions prevent the capture of content.

Site owners are encouraged to contact the Borthwick by [email or phone](#) with questions, concerns, or requests.

3.2 Data Protection

3.2.1 In order to mitigate the risk of capturing and exposing sensitive data, the Borthwick and its web archiving service provider have taken the following steps:

- The Borthwick does not knowingly capture sensitive or private content as part of its web archiving programme. Password-protected content is outside the scope of the web archiving programme.
- Prior to the capture of any web content, Borthwick staff consult with site owners to:
 - Ensure that we and site owners understand what types of content could be considered sensitive or confidential;
 - Assess whether the website in question is likely to contain any such material.
- Web archives are made publically accessible following a defined embargo period of at least six months.
- Site owners are asked to notify the Borthwick immediately if they become aware of any sensitive or infringing data that may have been published to their site and/or captured by our web crawlers. This is required regardless of whether a site is crawled once or at regularly scheduled intervals, such as semi-annual or quarterly. Site owners are responsible for communicating any such breaches throughout the full duration of their agreement with the Borthwick.

POLICY IMPLEMENTATION DOCUMENTS

4.1 This policy is supported by and should be read in conjunction with the following policy and implementation documents:

- Access to Collections policy
- Collections Care and Conservation policy
- Collections Information Policy
- Collections Management Policy
- Collections Development Policy
- Digital Preservation Policy
- University Archives Collection Policy
- University of York Information and Records Management Policy

OVERSIGHT

5.1 The Digital Preservation Archivist is responsible for overseeing policy implementation.

Document history and status

May 2023	D Waugh	Completed first draft
Jul 2023	D Waugh	Revised based on feedback provided by G Brannan, C Fonge, D Burt

Aug 2023	D Waugh	Approved
----------	---------	----------

Review cycle 5 years
Date of next review August 2028