

Queens Borough Public Library Web Archives Collection Development Policy

Introduction:

Queens Borough Public Library's traditional functions of acquiring, cataloging, preserving and serving collection materials of historical importance to the people of Queens extend to digital materials, including web sites. The Library uses web archiving to acquire and make permanently accessible, born digital works that are playing an increasingly important role in the intellectual, governmental, commercial and creative life of the Borough of Queens. These items include but are not limited to websites, social media, online audio video streaming, blogs, online media publications, and static document available online such as pdfs and jpegs. Web archiving preserves as much of the web-based user experience as technologically possible in order to provide future users accurate snapshots of what particular organizations and individuals presented on the archived sites at particular moments in time, including how the intellectual content (such as text) is framed by the web site implementation.

The Queens Borough Public Library Web Archives is available through the web archiving service, Archive-it.org, which incorporates a version of the open source crawling software Heritrix. Archive-it.org is an affiliate of the nonprofit organization, Internet Archive. Depending on collection guidelines and the nature of individual websites, websites may be recaptured at regularly scheduled intervals, such as semi-annual or quarterly. The Queens Borough Public Library Collection can be located at <https://archive-it.org/home/queenslibrary>.

This document will function as a written guideline for the continuing acquisition and capture of web materials that will inform researchers, community members and library patrons about the activities of the Queens Borough Public Library and life in the borough of Queens. In order for items to be added, they must provide a historically relevant picture of the geographic, cultural, economic, political and social life of Queens. This guide will aid in defining the scope and nature of this collection for patrons and other repositories, as well as our public services co-workers within the Queens Borough Public Library.

Scope of the Queens Borough Public Library Web Archive

The Web Archives will contain content relating to the history, culture, and current events of the Borough of Queens. Collections will build off of physical collections present in the Archives at Queens Library to fill holes in collection content and to add additional context. Metadata Services staff will archive websites requested by the Government Affairs and Communications Departments pertaining to the activities of the library and local government offices.

Content will cover the following areas:

- Residential life

Last revised 8/27/2018

- Geography
- Economic Activity
- Government Activities
- Business Activities
- Cultural Events
- Spontaneous Events
- Educational Institutions
- Local Art
- Local Community Groups/Organizations
- Social Movements/Grassroots Movements
- Race and Ethnicity
- Minorities such as LGBTQIA persons or groups
- Religious Groups
- Transportation
- Leisure
- Labor

Delivery platforms will include:

- General Websites
- Digital newspapers/journals
- Video
- Social media
- Blogs
- PDFs

Targeted Users:

The unique web materials of the Queens Borough Public Library Web Archive is made available to all people who have access to internet regardless of location. Users of the collection include those interested in the people, places, businesses/organizations, and events of Queens. Other users include library staff, students and educators, business persons, historic preservationists, journalists, and all other interested persons.

Collection Guidelines:

The following guidelines apply to all web content to be included in the Queens Borough Public Library Web Archive:

- Value – Web content selected have either evidential value (documentation of the functioning of an organization or individual) or research value and should maintain a complete and accurate representation of the original content to the extent possible. Web content that is unique or

extremely rare will be considered for accession even if they are out of the scope of the collection.

- Institutional Value- Web content that is created and released by the Queens Borough Public Library. This includes websites, social media, and blogs from all departments of the library system.
- Websites that complement and/or fill in gaps of existing collections or materials in the Archives at Queens Library.
- Websites with unique content that is periodically updated or possess ephemeral qualities.
- Web content that will not duplicate information that can be found in other web archiving collections

Intellectual Property Rights

Copyright protects the rights of the creators of published and unpublished original works. The copyright holder has exclusive rights to reproduce the work, prepare derivative works, distribute and/or sell copies of the original work and perform or display the work. The Queens Borough Public Library Web Archive will follow the section 108 Study Group's recommendations to the Copyright Act for website preservation (<http://www.section108.gov/docs/Sec108StudyGroupReport.pdf>). According to this report, archives and libraries have the right to crawl content that is publically available on the web that is not password protected, protected by robots.txt, and does not require a subscription. All governmental websites should be freely available to web crawlers. For inquiries about content captured and preserved by the library, including requests for removal of certain material, please see our policy: <http://digitalarchives.queenslibrary.org/Legal>.

In order to ensure the rights of the property owners are respected, Queens Borough Public Library with follow the follow guidelines:

- All robots.txt exclusions, with the exception of those used for social media, will be respected. Queens Library will only capture public posts on social media outlets.
- If QBPL Web Archive wishes to pursue a crawl that utilizes robots.txt outside the context of social media content, it will contact the content creator to negotiate access to the site.