QUEER JAPAN
WEB ARCHIVE

Ivy Plus Library Confederation, Japan LGBTQ Web Archiving Group
Collecting Ephemera
The Initial Members

Duke University
   Kristina Troost

Harvard University
   Mariko Honshuku
   Kuniko Yamada McVey

Princeton University
   Setsuko Noguchi

Yale University
   Haruko Nakamura

* Columbia University
   Sachie Noguchi
WHAT IS IVY-PLUS?

Ivy Plus Libraries Confederation Web Resources Collection Program

Ivy Plus Libraries Confederation

The Ivy Plus Libraries Confederation is a partnership between thirteen leading academic research libraries — Brown University, the University of Chicago, Columbia University, Cornell University, Dartmouth College, Duke University, Harvard University, Johns Hopkins University, the Massachusetts Institute of Technology, the University of Pennsylvania, Princeton University, Stanford University and Yale University — that collectively provide access to a rich and unique record of human thought and creativity through resource sharing and collaboration.

WEB COLLECTING

The Ivy Plus Libraries Confederation Web Collecting Program is a collaborative collection development effort to build curated, thematic collections of freely available, but at risk, web content in order to support research at participating Libraries and beyond. Participating Libraries are: Brown, Chicago, Columbia, Cornell, Dartmouth, Duke, Harvard, Johns Hopkins, Penn, Princeton, and Yale.

COLLABORATIVE COLLECTIONS

FAQ for Website Owners

COLLABORATIVE ARCHITECTURE, URBANISM AND SUSTAINABILITY WEB ARCHIVE (CAUSEWAY)

Browse/Search CAUSEWAY

- Example
  - Contemporary Composer Web Archiving (927 websites)

- Latin America and Caribbean Contemporary Art Web Archiving (76 websites)
WHAT IS IVY-PLUS?

Ivy Plus Libraries Confederation Web Resources Collection Program

IVY PLUS LIBRARIES CONFEDERATION

The Ivy Plus Libraries Confederation is a partnership between thirteen leading academic research libraries — Brown University, the University of Chicago, Columbia University, Cornell University, Dartmouth College, Duke University, Harvard University, Johns Hopkins University, the Massachusetts Institute of Technology, the University of Pennsylvania, Princeton University, Stanford University and Yale University — that collectively provide access to a rich and unique record of human thought and creativity through resource sharing and collaboration.

WEB COLLECTING

The Ivy Plus Libraries Confederation Web Collecting Program is a collaborative collection development effort to build curated, thematic collections of freely available, but at risk, web content in order to support research at participating Libraries and beyond. Participating Libraries are: Brown, Chicago, Columbia, Cornell, Dartmouth, Duke, Harvard, Johns Hopkins, Penn, Princeton, and Yale.

COLLABORATIVE COLLECTIONS

FAQ for Website Owners

COLLABORATIVE ARCHITECTURE, URBANISM AND SUSTAINABILITY WEB ARCHIVE (CAUSEWAY)

Browse/Search CAUSEWAY

- Budget
- Web Resource Collection Librarian
- Bibliographic assistance
- Subscription to AI account (2 tera byte)
- Program Supports
- Maintaining AI account
- Catalog records of collections
- Sends notifications
- Consultations
- Division of work
  - Scope and description
  - identifying websites
  - Provide metadata
Website Appraisal Checklist

The Website Appraisal Checklist is a tool to use when determining whether the website has permanent research value, and whether or not it should be a part of your collection.

Content of the website

1. What does the website document?
2. What value does the website have?
   - Types of values are:
     - Evidential Value: The quality of records that provides information about the origins, functions, and activities of its creator.
     - Informational Value: The usefulness or significance of materials based on their content, independent of any intrinsic or evidential value.
     - Intrinsically-Intrinsic Value: The usefulness or significance of an item derived from its physical or associational qualities, inherent in its original form and generally independent of its content, that are integral to its material nature and would be lost in reproduction. Visit the National Archives and Records Administration website for examples of Intrinsic Values:

3. Is the content of the website duplicated elsewhere in your collection? Is the content duplicated in another library or archival repository?
4. What use do you anticipate for the archived website?
5. Does the content of the website suggest how often it should be crawled? For example, a calendar of events usually changes on a monthly basis and may need to be crawled monthly.

Organizational mission and collection strategy

6. How does the website fit the goals of the University and its mission?
7. What is the relationship of the website to your current collection development policies or collecting scope?

Legal issues

8. What is the provenance of the website? Does the website contain vital records?
   - Vital records are defined as recorded information that must be protected in the event of an emergency or disaster because of severe consequences to the office and the University as a whole if the record is lost or destroyed.
   - Depending on use agreement. Does the effort of determining the ownership of the content justify collection?

9. How do legal guidelines apply to the website?
   - Please refer to the documents on copyright/permission of web archive
10. What restrictions apply to the website?
    - Does the website contain any sensitive or embargoed records? Typically archived websites are open to the public.
11. Does the information contained on a website fall under the review of a records retention schedule? Would this information be characterized as “Non-Permanent” in value to an organization and therefore be scheduled for deletion after a retention period has expired?

Technical issues

12. What is the storage medium? Is the information documented elsewhere in an easier to use format?
13. What are the costs of acquiring and preserving the website?
    - Will your time and resources be used efficiently through an appraisal of web content?
    - How much processing time will be involved?
14. Does an archived website need to function as it did as a live site on the web?
    - Does it use streaming video? Many web crawlers cannot crawl streaming AV or dynamically produced content.
15. Do any sites/pages require authentication (password protected)? Web crawlers will not capture content that requires authentication.
16. Do you want downloadable media archived? Do you have a large number of files to be archived?
17. Do your web content use JavaScript? JavaScript calls to external libraries may prove to be problematic. There is a possibility that some JavaScript plus CSS features in websites will not be archived.
18. Is your site database driven? These would include forms and searching capabilities that are difficult to archive.
19. It may be necessary to determine whether the site once existed under a different domain name if it was part of an online service like America Online (AOL) that may no longer exist. Visit [www.site.com](http://www.site.com) to learn more.
THE MEMBERS

Duke University
  Kristina Troost
Harvard University
  Mariko Honshuku
  Kuniko Yamada McVey
Princeton University
  Setsuko Noguchi
Yale University
  Haruko Nakamura

Yoshie Yanagihara 柳原良江
Assistant professor, Tokyo Denki University

Tetsuyuki Shida 志田哲之
Part time lecture, Waseda University

Samantha Abrams
Web Resources Collection Librarian, Ivy Plus Libraries Confederation
SETTING UP

- Determine the Scope

- Experience of collecting archives

- Selections
  - Selection list and Scope size
  - Duplicates and exclusions

公共机关・自治体 Public agencies / municipalities
公共助成金・基金 The Foundation for Aids Prevention 額に係する研究

啓発・支援・自助
Enlightenment / support / self-help

パートナーシップ・家族 Partnership / Family

活動家 Activists

大学・研究機関 Universities / research institutions

公共機関・自治体 Public agencies / municipalities

政治家 / politicians
Determine the Scope
- Experience of collecting archives

Selections
- Scope size and selection list
- Duplicates and exclusions
  - NLA - Japan : Sex and gender : LGBTQ
  - NDL - WARP
  - Universities and Independent Administrative Institutions.
- Aggregates
Setting Scope

- Duplications from Internet Archives
- In-depth crawling (video, pdf, pictures)
- Narrow curations
- Better metadata
- Local digital preservation
- SNS – Twitter, Facebook, Instagram
Description

Queer Japan Web Archive is a joint initiative by librarians for Japanese Studies at Duke, Harvard, Princeton, and Yale universities under the auspices of the Ivy Plus Libraries partnership. This project aims to preserve websites created by non-profit Japanese organizations which research, advocate for, support and provide services to LGBTQ communities. The non-profit organizations provide social services for the LGBTQ community, such as quality of life services, real estate, employment, educational assistance, helplines, workshops, mental health services, information about sexually transmitted diseases such as HIV/AIDS, and legal information. By archiving these websites, the group strives to guarantee that these potentially ephemeral websites are preserved and remain browsable by researchers and scholars.

Queer Japan Web Archive は、Ivy Plus Librariesのパートナーシップを得て、デューク、ハーバード、プリンストン、イェール各大学の日本研究ライブラリアンによって運営される共同イニシアチブです。このプロジェクトは、日本の非営利団体によるLGBTQコミュニティの研究、支援、サービス提供を目的として作成されたウェブサイトのアーカイブを目指しています。対象としている非営利団体が行っているLGBTQコミュニティの支援サービスには、生活の質向上、不動産、雇用、教育援助、ヘルプライン、ワークショップ、メンタルヘルス、HIV / AIDSのような性感染症、そして法律関係の情報などが含まれます。私たちは、研究者のために、こうした重要でありながら失われやすい情報への継続的アクセスの保全に努めます。
<table>
<thead>
<tr>
<th>網上のグループ</th>
<th>ウェブサイト</th>
<th>ウェブサイトタイトル</th>
<th>関連タイトル</th>
<th>公式メイカー</th>
<th>説明（仮想的）</th>
<th>日本語説明</th>
<th>エンゲージメント</th>
<th>ストックデスク（仮想的）</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stonewall Japan</td>
<td><a href="https://stonewalljapan.org/">https://stonewalljapan.org/</a></td>
<td>Stonewall Japan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Business networks;</td>
<td></td>
</tr>
<tr>
<td>Rebit</td>
<td><a href="https://rebitjapan.org/">https://rebitjapan.org/</a></td>
<td>りびっと</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Stobewall Japan (Organisation);</td>
<td></td>
</tr>
<tr>
<td>Proud Life</td>
<td><a href="http://proudlife.org/">http://proudlife.org/</a></td>
<td>プラドライフ～名古屋から多様性の発信を～</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Stobewall Japan (Organisation);</td>
<td></td>
</tr>
<tr>
<td>Loud: Lesbians of Undesirable Drive</td>
<td><a href="http://space-loud.org/">http://space-loud.org/</a></td>
<td>Loud</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Stobewall Japan (Organisation);</td>
<td></td>
</tr>
<tr>
<td>LGBTの家族と友人をつなぐ会</td>
<td><a href="http://lgbt-family.or.jp/">http://lgbt-family.or.jp/</a></td>
<td>LGBTの家族と友人をつなぐ会</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Stobewall Japan (Organisation);</td>
<td></td>
</tr>
</tbody>
</table>
PLANNING FOR THE FUTURE: HOME PAGE

- Scope & Descriptions in both J and E
- FAQ
  - E.g. Snapshot of Japan 2016-2018 (Stanford)
- Links to related collection as well as FB, Twitter, Instagram, etc.
- Forms for suggesting websites
- Featured websites
  - E.g. Human Rights Web Archives (Columbia)
- Secondary curated contents
  - E.g. Timeline: Japan’s Constitutional Discussion (Harvard)
Data Mining and Data Manipulations

What sort of formats we will collect?
(XML, HTML, PDF, Doc w/o links/TXT/TEI)

How are we going to organize data?

APIs? (in order to retrieve certain data)

OCR for PDFs?

Cloud source? or Scraping metadata from websites?

State Election Website
THANK YOU!