What Is Worth Saving?
Selection and Curation in Web Archiving

Lori Donovan, Internet Archive
Tessa Fallon, Columbia University
Geoffrey Harder, University of Alberta
Web Archiving at Columbia University Libraries

Tessa Fallon, Web Collection Curator
What Is Worth Saving? Selection and Curation in Web Archives

Print/continuing/digital resource selection

All CUL web archives are:
- Publicly available
- Harvested using Archive-It
- Accessible via CLIO (OPAC), WorldCat, and Archive-It

Selection models by collection:
- Human Rights Web Archive
- Avery
- Burke
- RBML
- University Archives
What Is Worth Saving? Selection and Curation in Web Archives

- Developed as focal collection of grant-funded project
- Proposed scope:
  - at-risk human rights NGOs
  - Site-level collection (entire sites instead of selected portions)
  - Global south
  - Locally or nationally focused
- Initial collection of sites:
  - Subject specialists
  - Project staff
  - Delicious social bookmarking

Challenges:
- Fluidity of organization missions
- Defining “at-risk”
- Artifact vs. resources
Scope expanded to include: NHRIs, truth commissions, tribunals, Amnesty International chapters

Survey: 2011

- Graduate students with expertise in: Russian language, Sub-Saharan Africa, and Latin America
- Guidelines: by region/country/topic/organizations
- Ranking
- Consultation with Subject Specialist
  - Prioritizing: underrepresented countries, topic of particular importance to a country, unique resources, HR-dedicated
What Is Worth Saving? Selection and Curation in Web Archives

Collection

Selection

Acquisition

HRWA Collection Development Overview

SURVEY
Broad criteria

Subject Specialists
Curators
Students

VETTING
Refined criteria

Subject Specialists
Curators

Curators

iPRES 2011 Singapore
Role of the Curator

Curator Evaluation: topical and/or technical

Collection Statistics:
- Crawled sites
- Live seeds
- Languages
- Countries
- Subjects
Selection: Players and Processes

- Selectors
  - Complement other collecting
  - Subject expertise

- Students
  - Structured survey
  - Consultation with subject specialists

- Curators
  - Subject expertise
  - Technical expertise
  - Knowledge of existing collections
Avery Fine Arts and Architecture

**Selector**
- Subject specialists
- Within collecting scope of Avery Library
- Submit IRCR

**Catalogers**
- Receive IRCR
- Complete bibliographic record
- Flag record for curator

**Curator**
- Requests permission from site owner
- Archives site
- Performs quality control
- Creates additional metadata

What Is Worth Saving? Selection and Curation in Web Archives

iPRES 2011 Singapore
What Is Worth Saving? Selection and Curation in Web Archives

Selector
- Collection Services Librarian
- Proposes thematic collection
- Works with curator to create collection guidelines
- Defines scope

Intern & Student Assistant
- Use guidelines to compile lists of sites
- Identify resources for additional site collection
- Tools:
  - Forms
  - Spreadsheets

Curator
- Requests permission from site owner
- Archives site
- Performs quality control
- Creates additional metadata

iPRES 2011 Singapore
University Archives
- Collecting to support University Archives’ mission, institutional history
- Capturing entire columbia.edu with focused quality control for areas prioritized by University Archives

RBML
- Collecting to supplement existing archives:
  - Erika Jong papers
  - Amiri Baraka papers
  - Carnegie Council
<table>
<thead>
<tr>
<th>Process</th>
<th>Human Rights</th>
<th>Avery</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Scope</strong></td>
<td>Human rights organizations</td>
<td>Historic preservation, urban planning</td>
</tr>
<tr>
<td><strong>Selectors</strong></td>
<td>Area Studies Subject Specialists, Students, Curators</td>
<td>Avery Subject Specialists</td>
</tr>
<tr>
<td><strong>Criteria for Collection</strong></td>
<td>Subject relevance, org. type, unique publications</td>
<td>Subject relevance</td>
</tr>
<tr>
<td><strong>Tools</strong></td>
<td>IRCR, public nomination forms, Google forms</td>
<td>IRCR</td>
</tr>
<tr>
<td><strong>Print/Other Collections</strong></td>
<td>CHRDR archives collections</td>
<td>NYC Landmarks Preservation, Commission Design. reports</td>
</tr>
</tbody>
</table>
Active collection vs. passive collection

Other possibilities:
- Collaborative web archiving
- Web archiving for datasets
- Single page web archiving
- Document-only web archiving

Dependence on resources:
- Staff
- Storage
- Budget
- Tools
- Collection plan
  - Project vs. program
  - Long term strategies
  - Finite or (seemingly) infinite number of relevant sites