

8th International Conference on Preservation of Digital Objects  
November 1-4, 2011, Singapore

# 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

# Selection at Columbia

- ✎ **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

# Human Rights Web Archive

- ⌘ 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

# Human Rights Web Archive (cont'd)

- ✎ **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

# HRWA Collection Development Overview

## Collection

SURVEY  
Broad  
criteria

Subject  
Specialists  
Curators  
Students

## Selection

VETTING  
Refined  
criteria

Subject  
Specialists  
Curators

## Acquisition

Curators

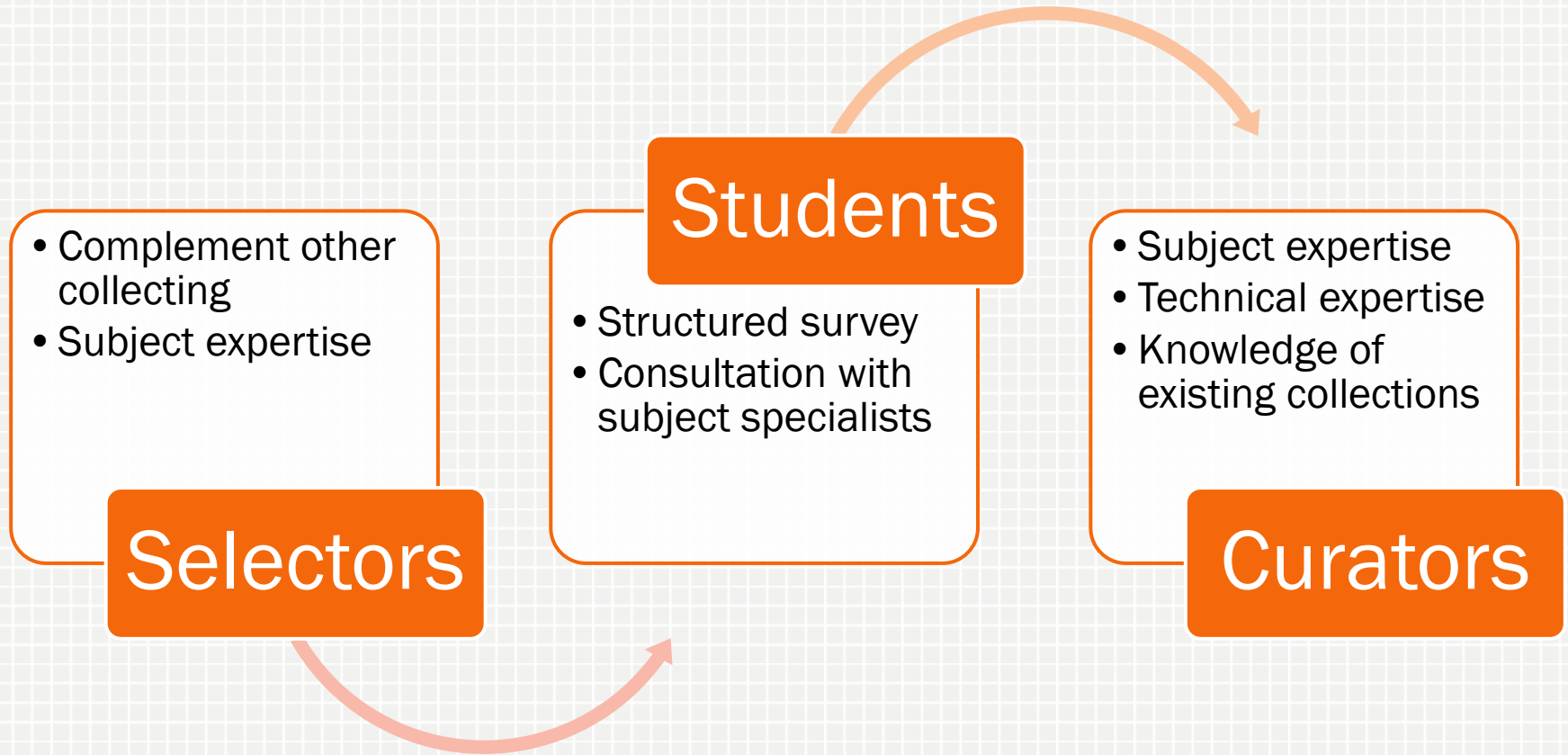
# Role of the Curator

☞ **Curator Evaluation: topical and/or technical**

☞ **Collection Statistics:**

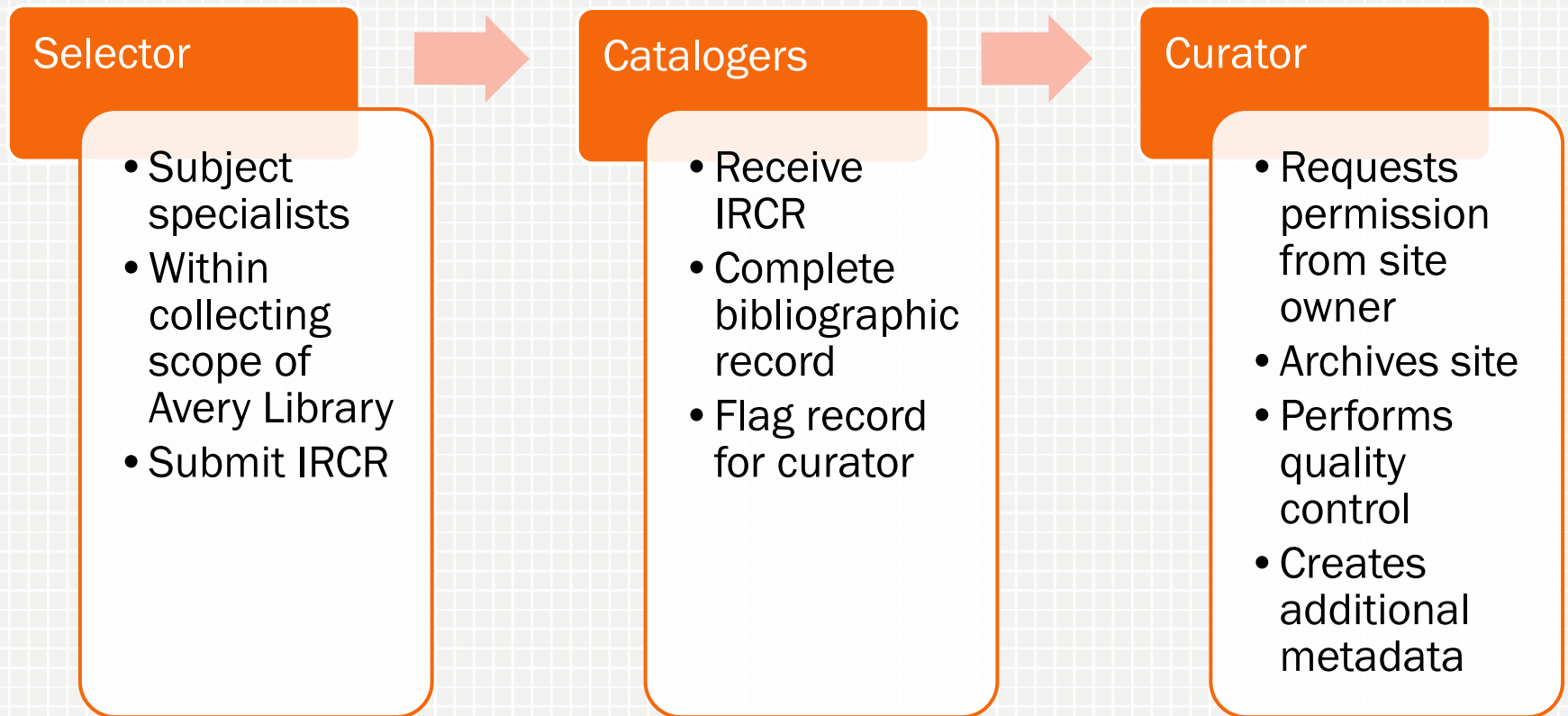
- **Crawled sites**
- **Live seeds**
- **Languages**
- **Countries**
- **Subjects**

# Selection: Players and Processes

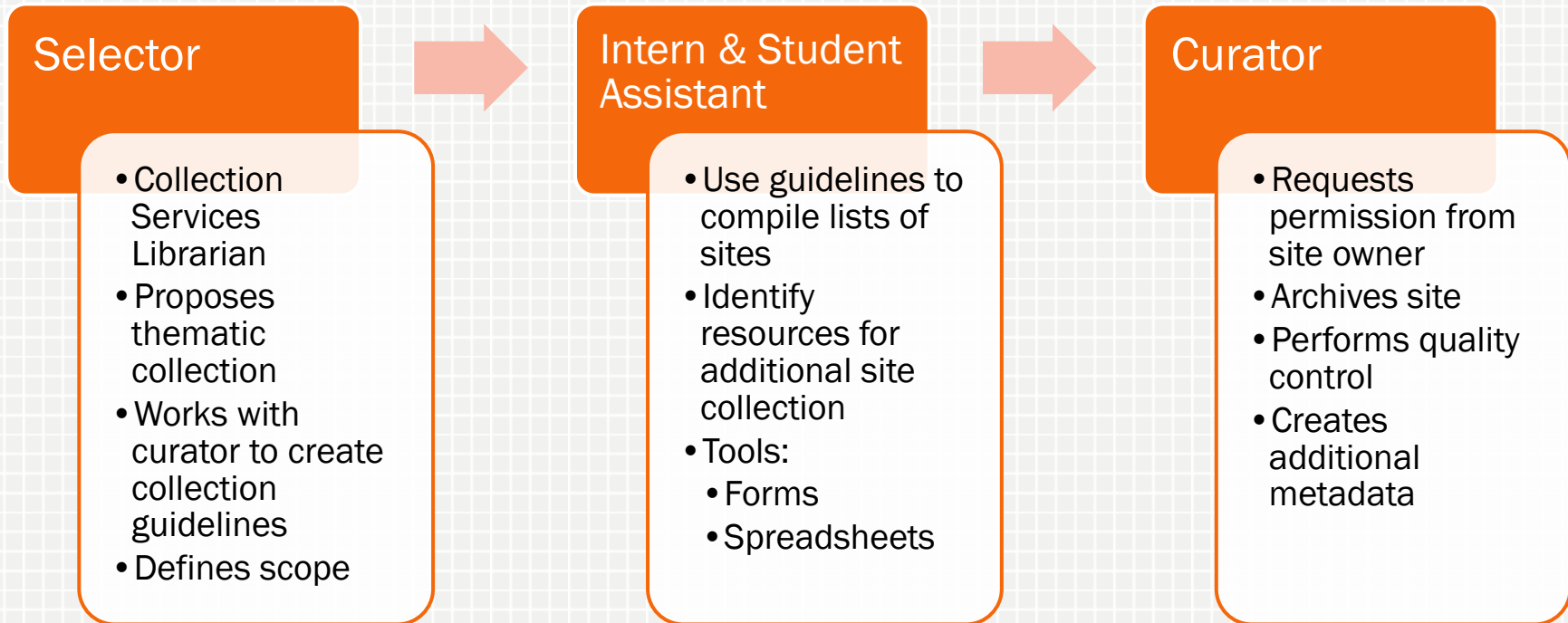




# Avery Fine Arts and Architecture



# Burke Theological Library



# Other collections...

## ☞ 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

## Process

SCOPE

SELECTORS

CRITERIA FOR COLLECTION

TOOLS

PRINT/OTHER COLLECTIONS

## Human Rights

Human rights organizations

Area Studies Subject  
Specialists, Students, Curators

Subject relevance, org. type,  
unique publications

IRCR, public nomination forms,  
Google forms

CHRDR archives collections

## Avery

Historic preservation , urban  
planning

Avery Subject Specialists

Subject relevance

IRCR

NYC Landmarks Preserv.  
Commission Design. reports

# University Archives

Columbia.edu domain

University Archives staff

Domain, UA staff designated priorities

Spreadsheets, domain assessment tools

University Archives

# RBML

Related to RBML archival holdings

RBML staff

Related to RBML archival holdings

Email

RBML Archives

# Burke

Religious institutions of NYC

Subject Specialist, Grad. Students, Curator

Religious tradition, location (NYC), neighborhood

Spreadsheets, maps

Burke archival collections

# Other considerations...

- ∞ **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