



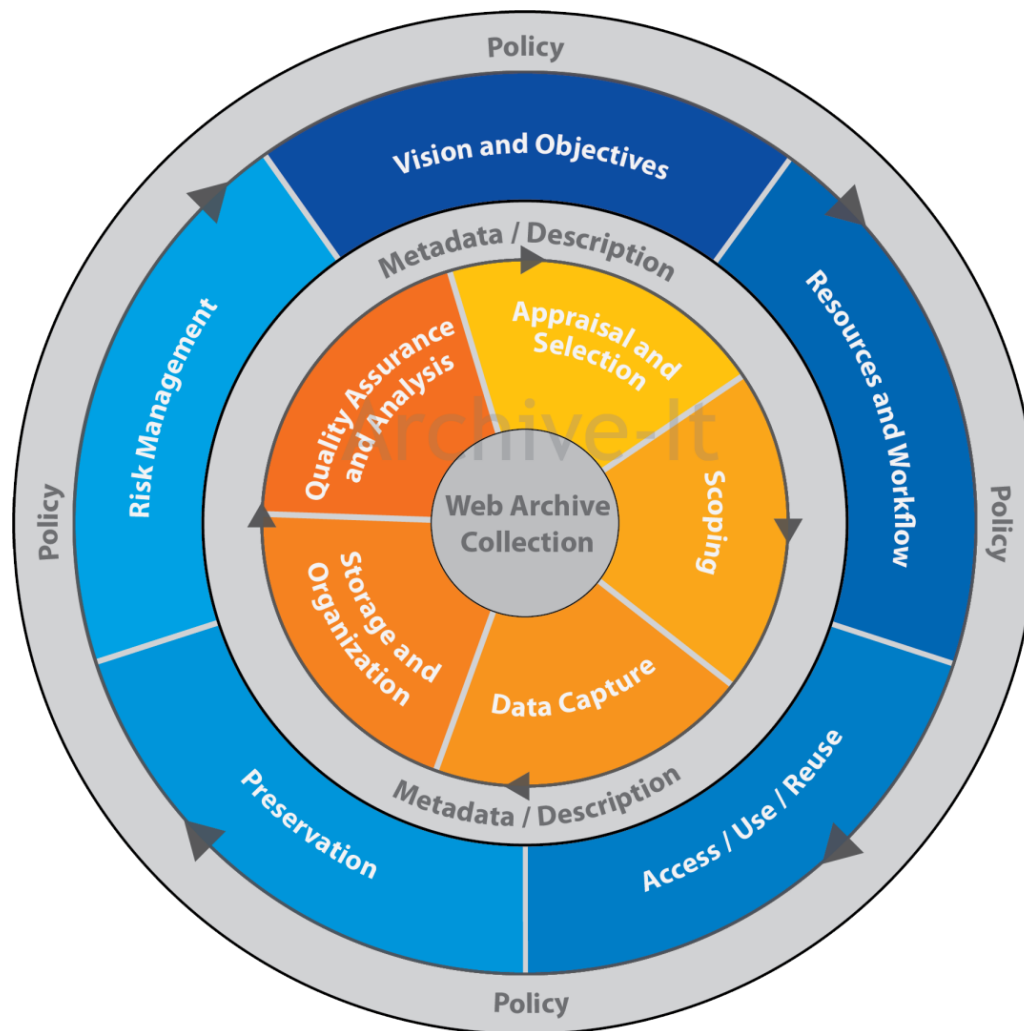
STANFORD UNIVERSITY LIBRARIES

Considerations for Strategic Web Archive Collection Development

Nicholas Taylor
Web Archiving Service Manager
[Stanford University Libraries](#)

[Curating Web Archives: Who Cares for Content?](#)
May 23, 2014

web archiving lifecycle

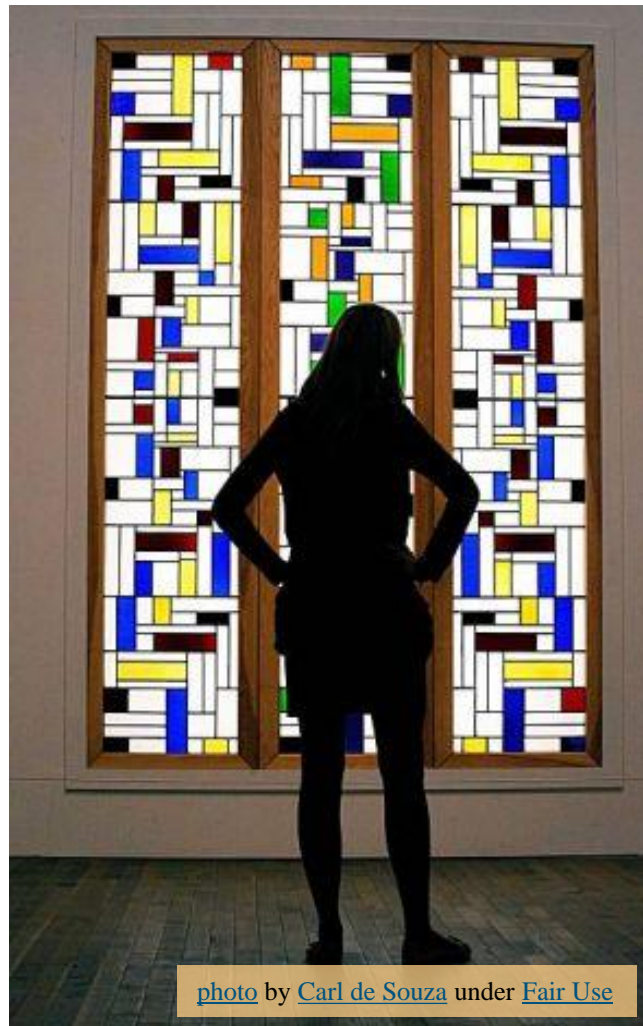




curator tools

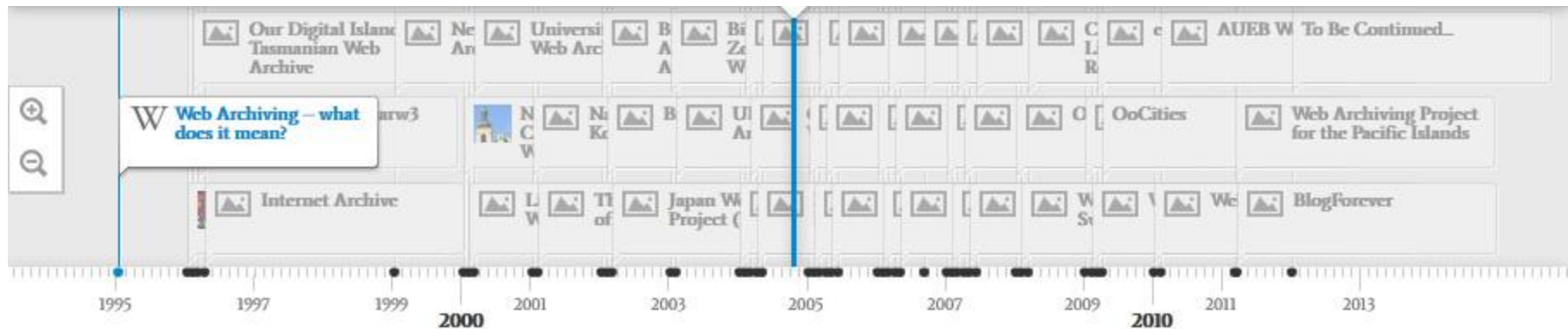
[illegible]

appraisal and selection



we are few

- **70** web archiving initiatives on Wikipedia
- **313** Archive-It partners
- **33** CDL WAS subscribing institutions



WebArchivists: "Timeline"

how much archived?



79%



16%



del.icio.us

68%

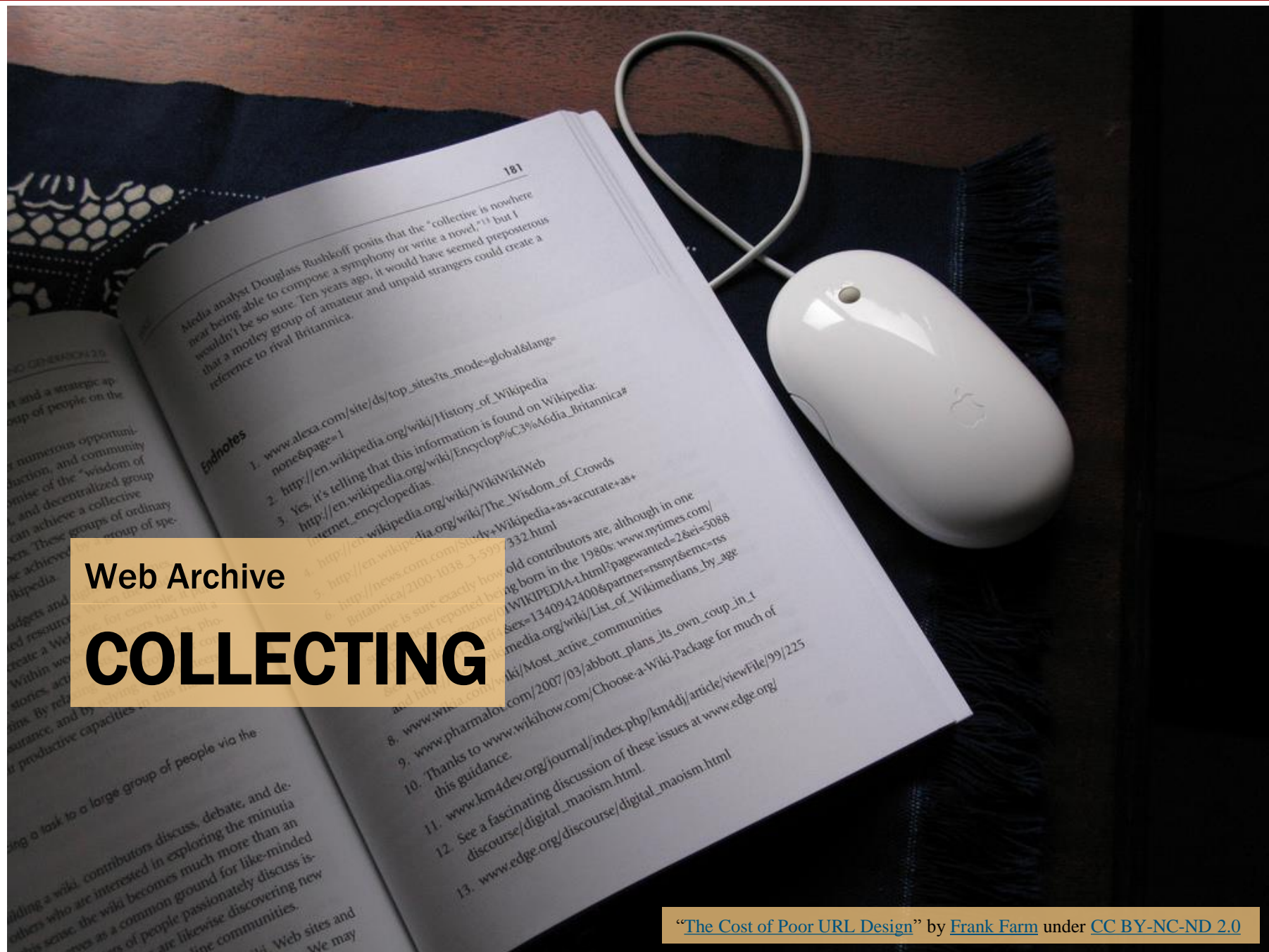


19%

selection determines preservation



“20130809-FS-LSC-0607” by U.S. Department of Agriculture under [CC BY 2.0](#)



Web Archive

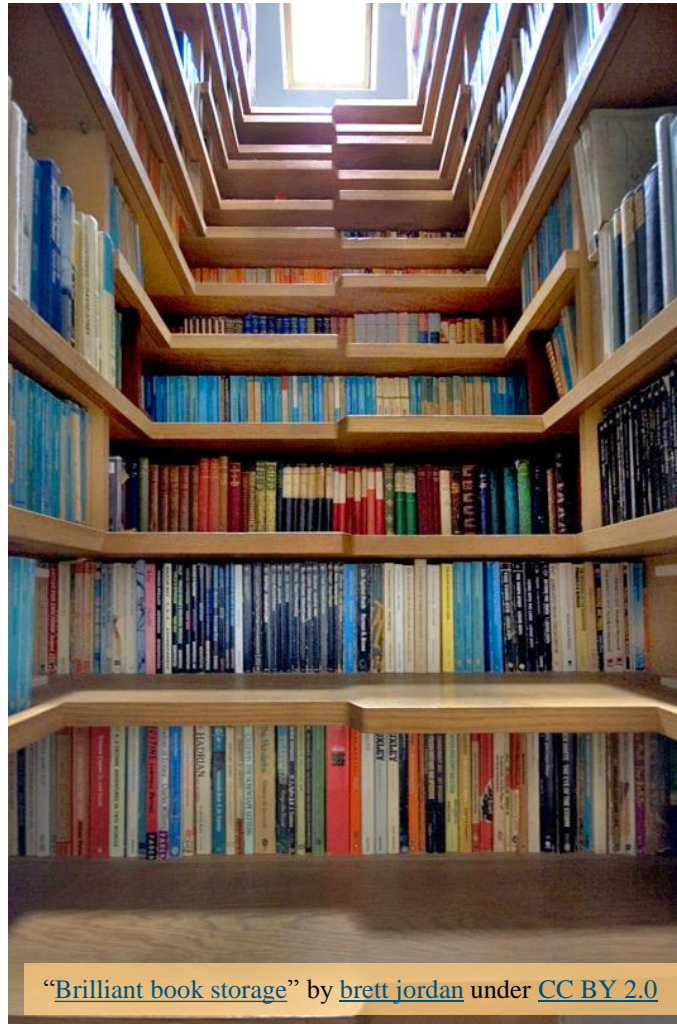
COLLECTING

subject expertise



Wordle: "People | Stanford University Libraries"

traditional collecting



“[Brilliant book storage](#)” by [brett jordan](#) under [CC BY 2.0](#)

collecting compared

traditional

- published
- one-time, up-front curation
- rivalrous, usable by a local service population
- comprehensive
- many copies
- purchase/license
- finite acquisition

web archives

- public
- ongoing curation
- non-rivalrous, potentially usable by anyone
- representative
- few copies
- permissioned/sanctioned
- contingent acquisition

how others collect



“2009 san diego comic-con: comics, still an elemental part of the con” by [george ruiz](#) under [CC BY 2.0](#)

necessary but not sufficient

- align with organizational **mission**
- support **research** and **teaching**
- preserve **institutional legacy**
- consider **history** and **geography**

necessary but not sufficient



*“In principle, the **collection development policy** for the Tamiment Library’s Web Archive **parallels** that of the Tamiment Library as a whole (labor and radicalism)”*

*In practice, this is complicated by (a) the **enormous size and variety of born digital materials** within Tamiment’s collecting scope...and (c) **resource restraints**. Thus the Library will not only have to **carefully appraise** materials, but to **set priorities and limitations**.”*

what not to collect



“War of the Worlds” by [7-how-7](#) under [CC BY-NC-ND 2.0](#)

sufficient-y

- collect within **subject area**
- focus on **at-risk** content
- collect content **previously collected in print**
- limit to particular **types of organizations**

sufficient?

- consider **what others are collecting**
- **don't** aim to **be comprehensive** (if you can't be)
- **complement** existing **strengths**
- prefer **current** and/or **unique** content
- mind **resource constraints**
- collect **publicly available** content
- anticipate **value to researchers**
- collect **content, not links** to content
- target specific **resource** or **format types**
- enable **designated research**

thank you!

