



STANFORD UNIVERSITY LIBRARIES

# Collection Development for Selective Web Archiving

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

[Archive-It Partner Meeting](#)  
August 2, 2016

# the Web is on fire



“Forest wildfire” by [Project LM](#) under [CC BY-NC-ND 2.0](#)

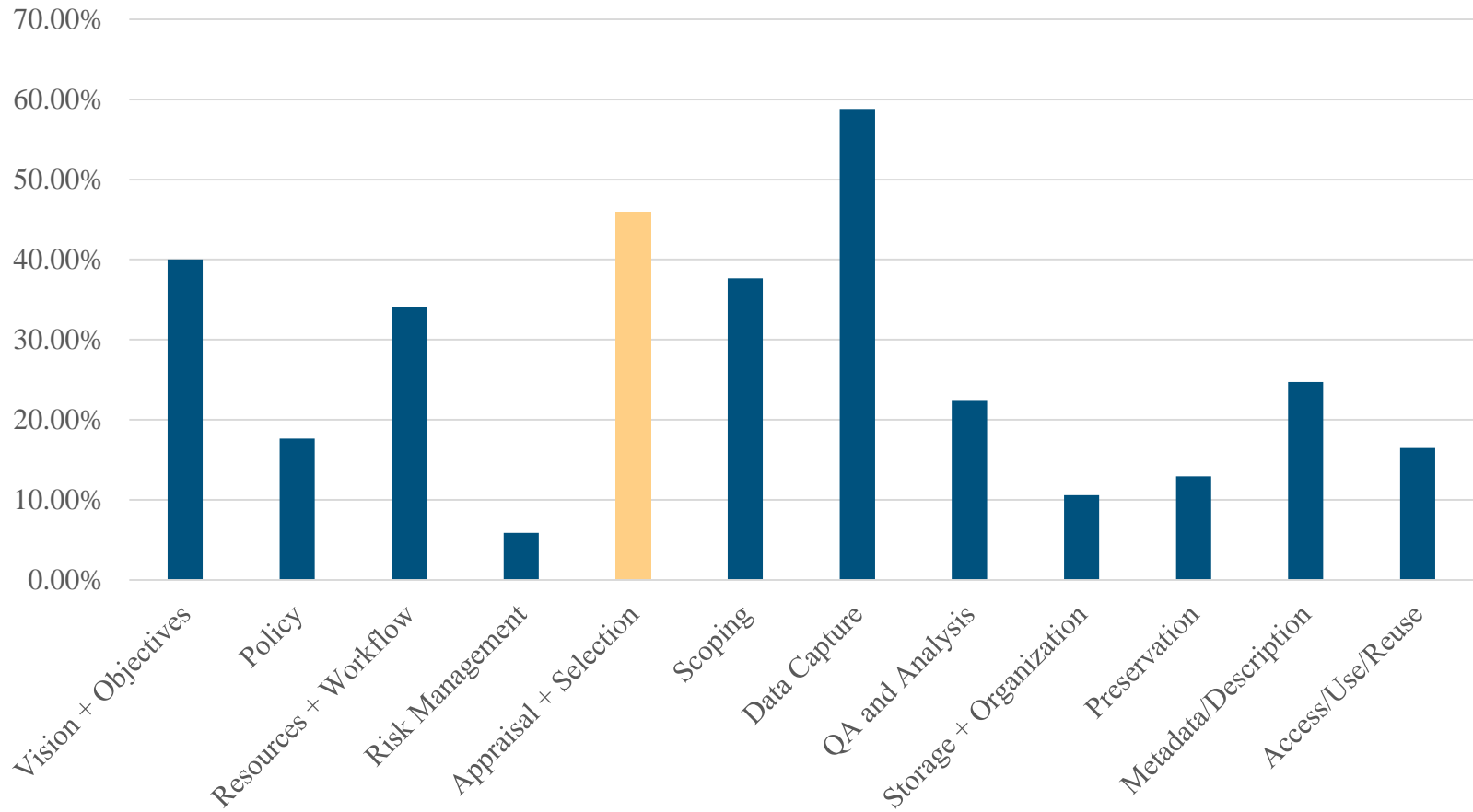
# what are we going to save?



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

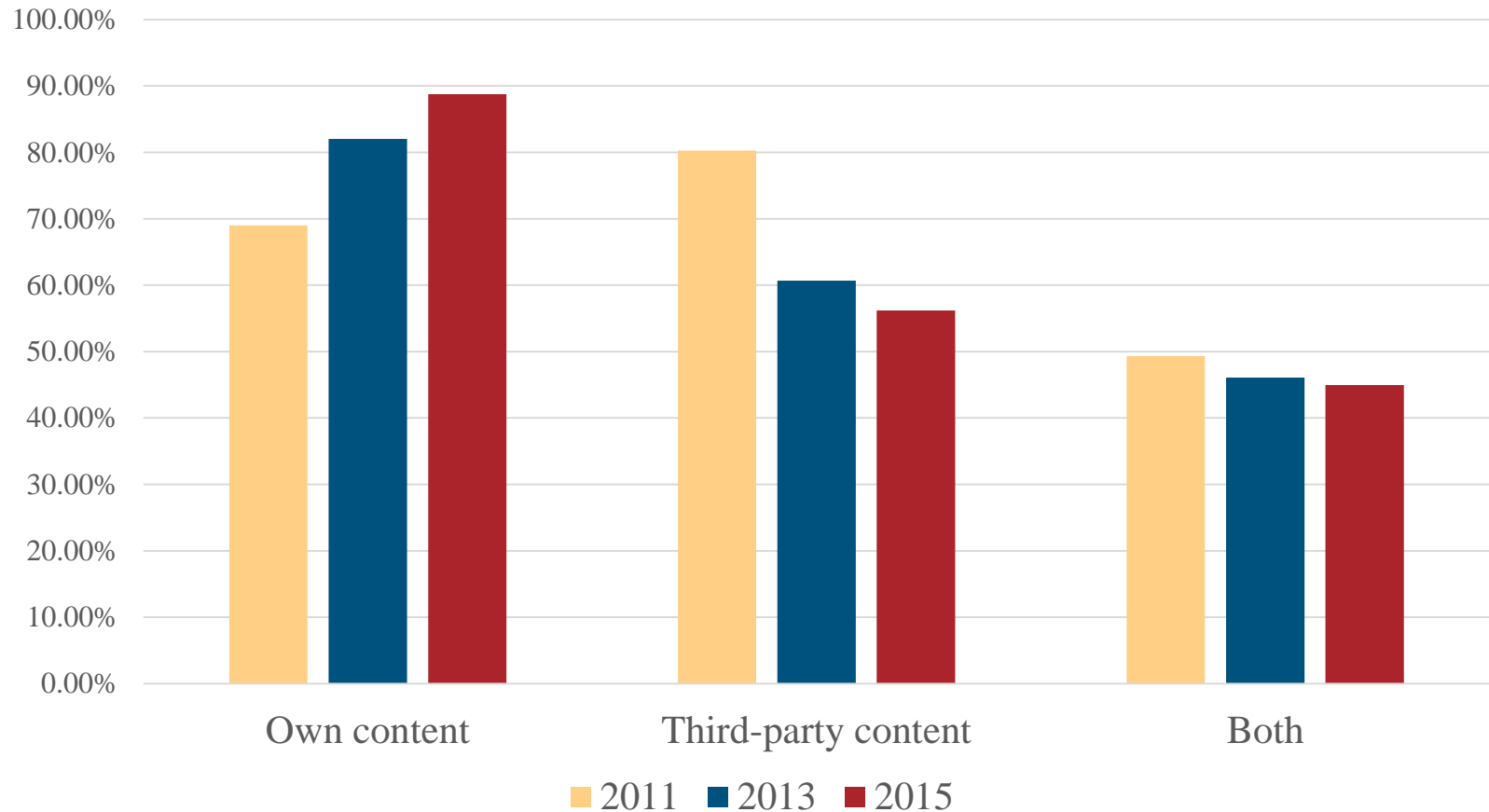


# an area of perceived progress

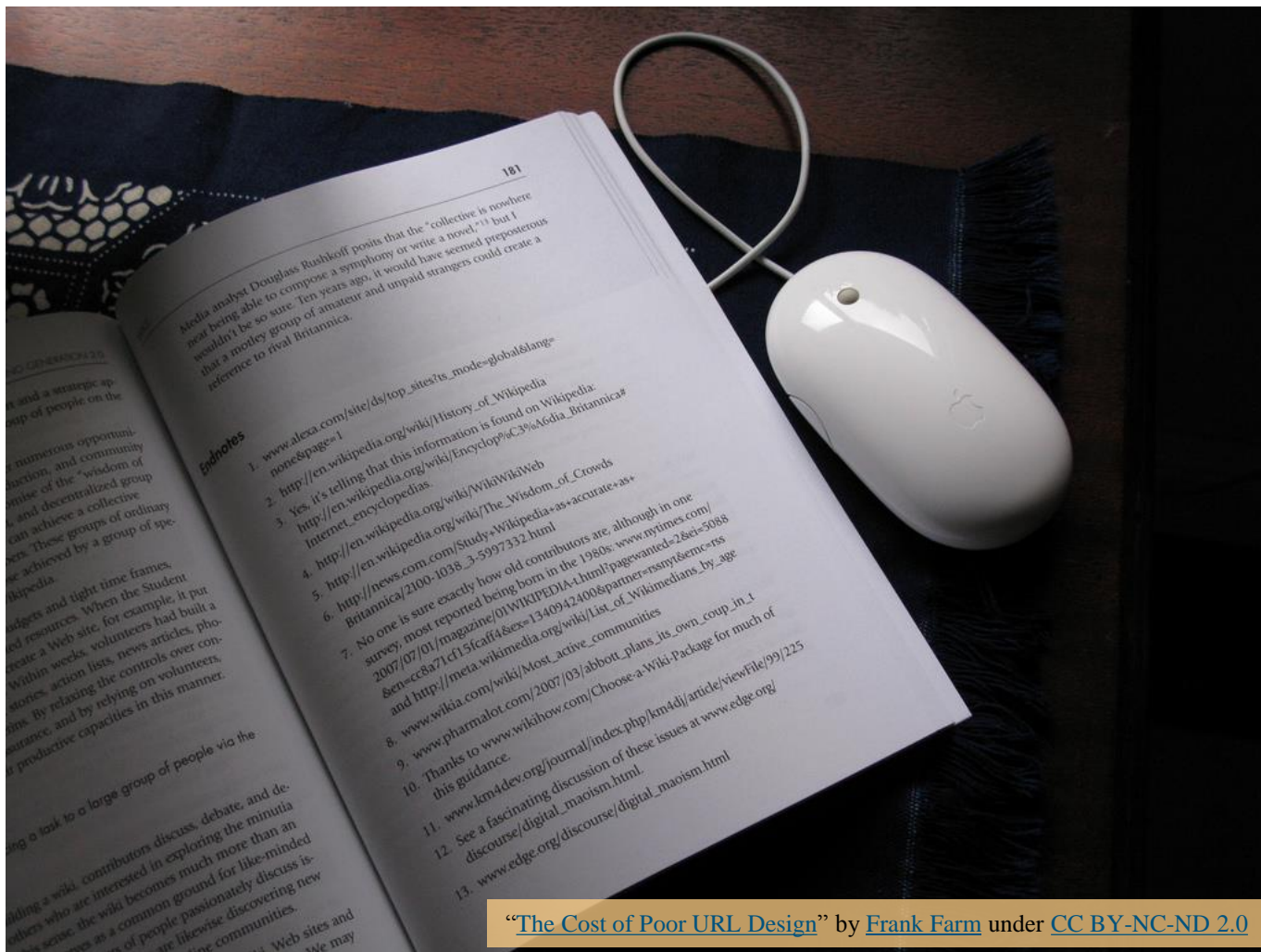


[NDSA](#): “2015 NDSA Web Archiving Survey”

# growth in archiving own content



# traditional + web content collecting



# subject expertise



Wordle: "People | Stanford University Libraries"

# 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**.”*



# focus on at-risk content





# complement collecting strengths





# observe resource constraints



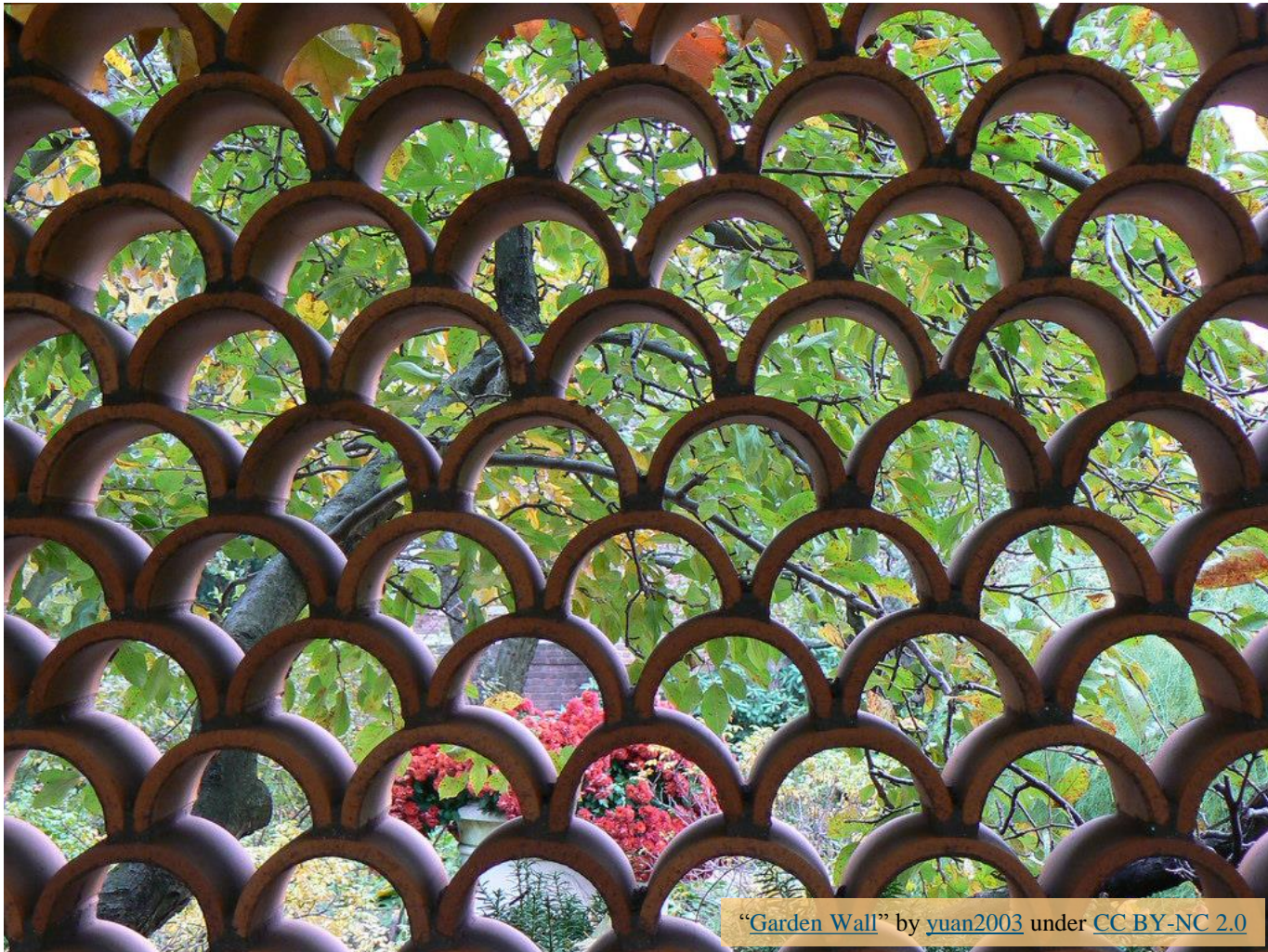
“[Abandoned](#)” by [Daniel D'Auria](#) under [CC BY-SA 2.0](#)

# consider what others are collecting





# consider others' access restrictions



“Garden Wall” by [yuan2003](#) under [CC BY-NC 2.0](#)

# assess value to researchers

October 2014 entities – Columbia U.'s Human Rights collection (~4M docs; 330GB)

## TOP 6 People Entities (of 1,887,523)



Barack Obama

Ilham Tohti

Julian Assange

George W. Bush

Pu Zhiqiang

Bashar al-Assad

## TOP 6 Organizations Entities (of 1,335,232)



## TOP 9 Locations Entities (of 373,155)



[Archive-It: "WANE Example Use Cases"](#)



# enable specific research

**L'ÈRE DU PEUPLE**

## La France insoumise jlm2017.fr

Cliquez ici pour appuyer la proposition de candidature de Jean-Luc Mélenchon

**À CHAUD**

Macron ahoho, la caravane passe jlm2017.fr

**À LA UNE**

La révolution citoyenne a déjà commencé - L'entretien de l'interview

Ensemble, construisons la France de demain

**VIDÉOS**

LE 28 AOÛT, ON POUVRIER ?

**VOUS AILLEURS**

NON ! À L'AÉROPORT

**À FOND**

PLAN B

**À FROID**

LA VILLE DE NICE

**LA FRANCE INSOUMISE**

La France insoumise jlm2017.fr

**MA CHAÎNE VIDÉO**

En 2016, nous avons encore le choix.

**ACTUALITÉS**

En 2016, nous avons encore le choix.

**JE SUIIS**

En 2016, nous avons encore le choix.

**En librairie**

La France insoumise

**Abonnez-vous au blog**

**Dossiers Inductuels**

**“Jean-Luc Mélenchon | Le blog”**

**ALAIN JUPPÉ 2017**

## « FACE AU TERRORISME, MA LIGNE EST CLAIRE : INTRANSIGEANCE ET SANG FROID. »

Propos recueillis par l'actu.fr

**Ensemble, construisons la France de demain**

**ACTUALITÉS**

En 2016, nous avons encore le choix.

**JE SUIIS**

**“Alain Juppé pour la France”**

**LA FRANCE POUR LA VIE**

## ON REPREND LA ROUTE REJOIGNEZ-NOUS !

Rejoignez nos équipes ENGAGEZ-VOUS

**ARTICLES**

Nicolas Sarkozy dénonce les orgues du gouvernement

**LA FRANCE POUR LA VIE !**

**https://www.sarkozy.fr/**

**MARINE 2017**

Bienvenue sur Marine2017.fr

**Marine Le Pen 2017**

**“Marine Le Pen 2017”**

# consider appropriate archiving tool



“Fruit Picker” by [Naoto Sato](#) under [CC BY-NC-SA 2.0](#)



# save content, not links



# prefer current, esoteric content



79%



16%



del.icio.us

68%



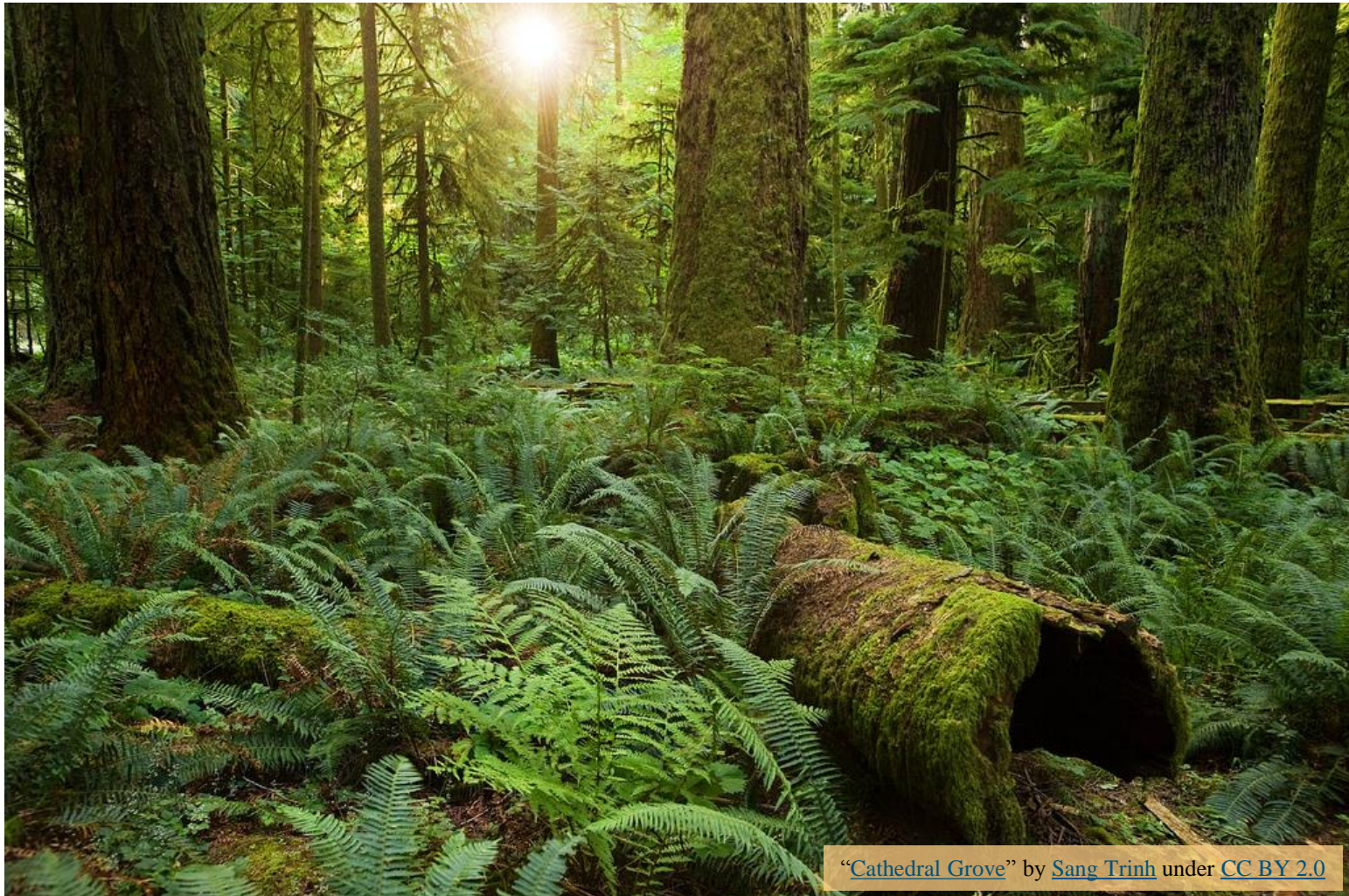
19%

# support community self-archiving





# together we can preserve the Web



“Cathedral Grove” by [Sang Trinh](#) under [CC BY 2.0](#)