STANFORD UNIVERSITY LIBRARIES

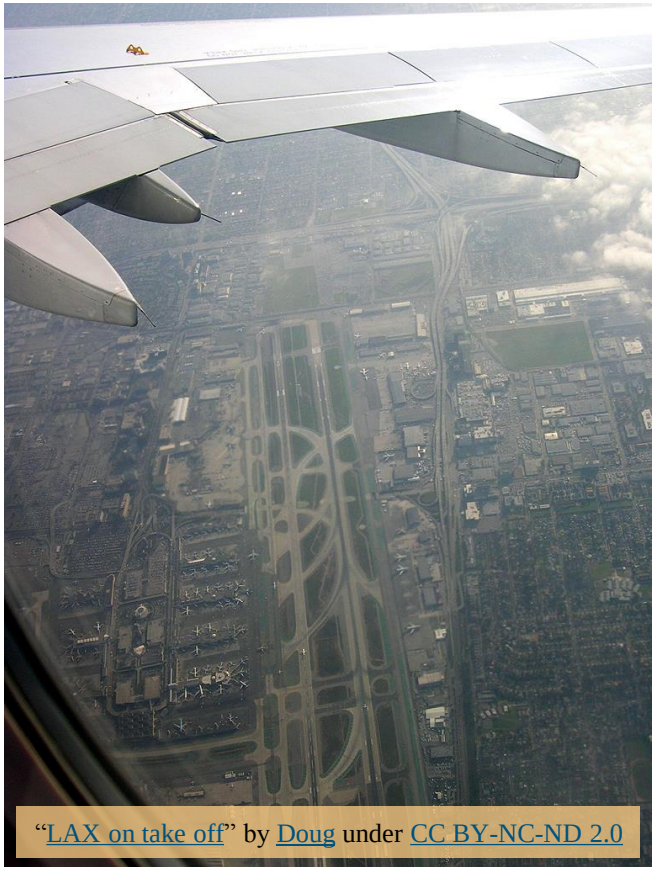
One Site Among Many: Stanford and Collaborative Technical Development for Web Archiving

Nicholas Taylor
Web Archiving Service Manager
Stanford University Libraries

PASIG 2016
March 11, 2016

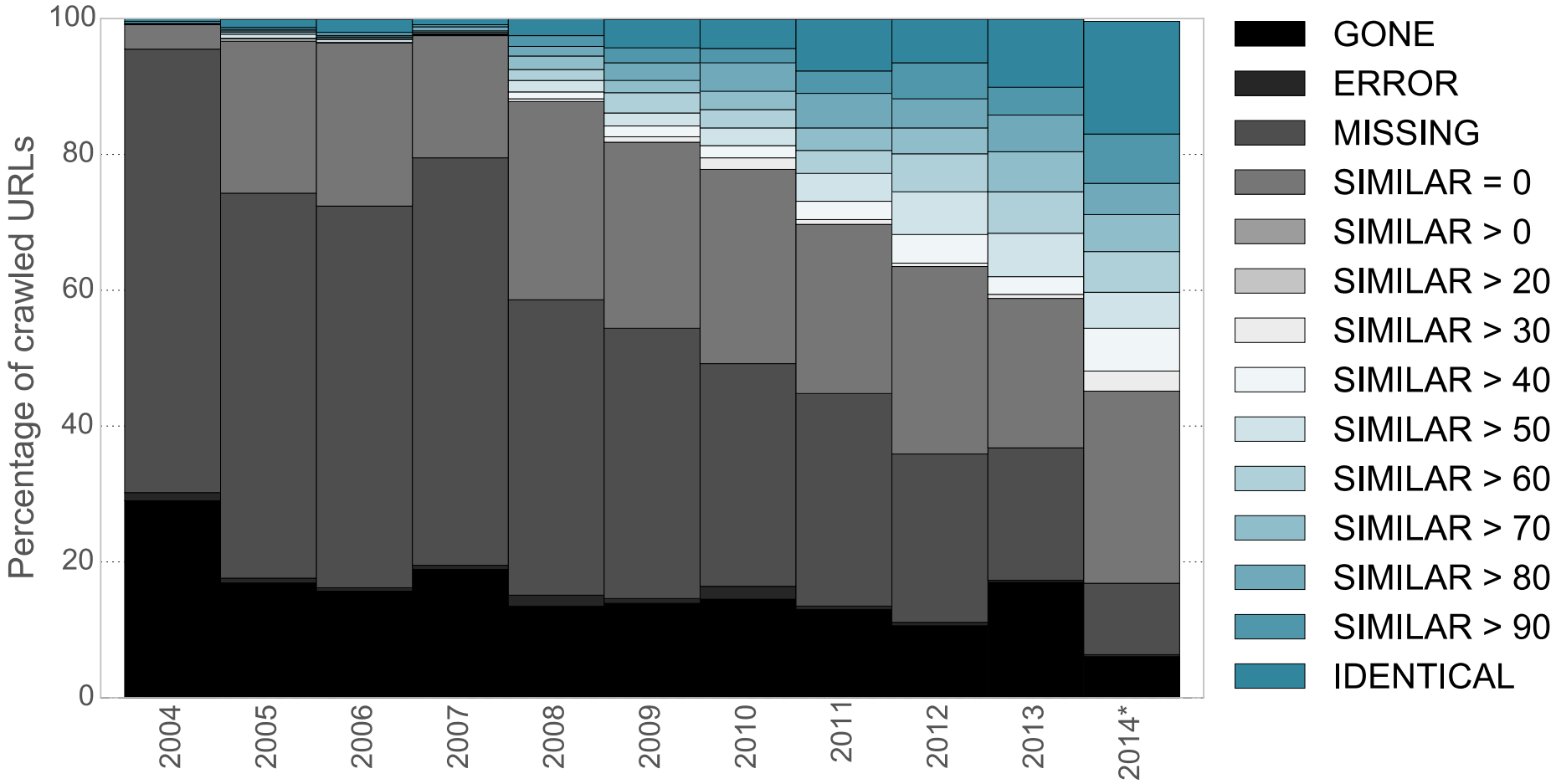
overview

- web archiving opportunity gaps
- situation of SUL web archiving
- APIs + community (technical) development



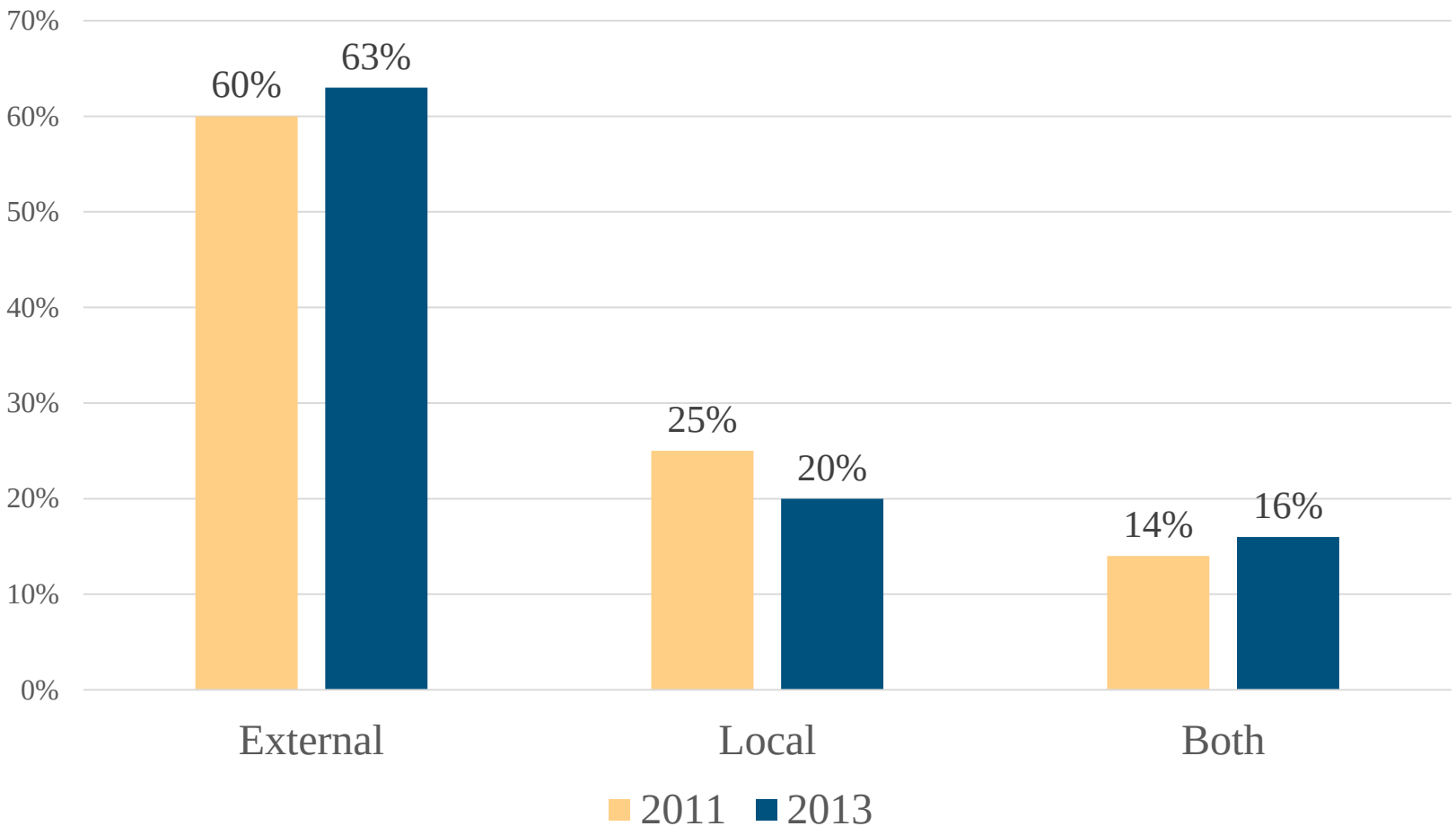
web content > preserved web content

link rot + content drift



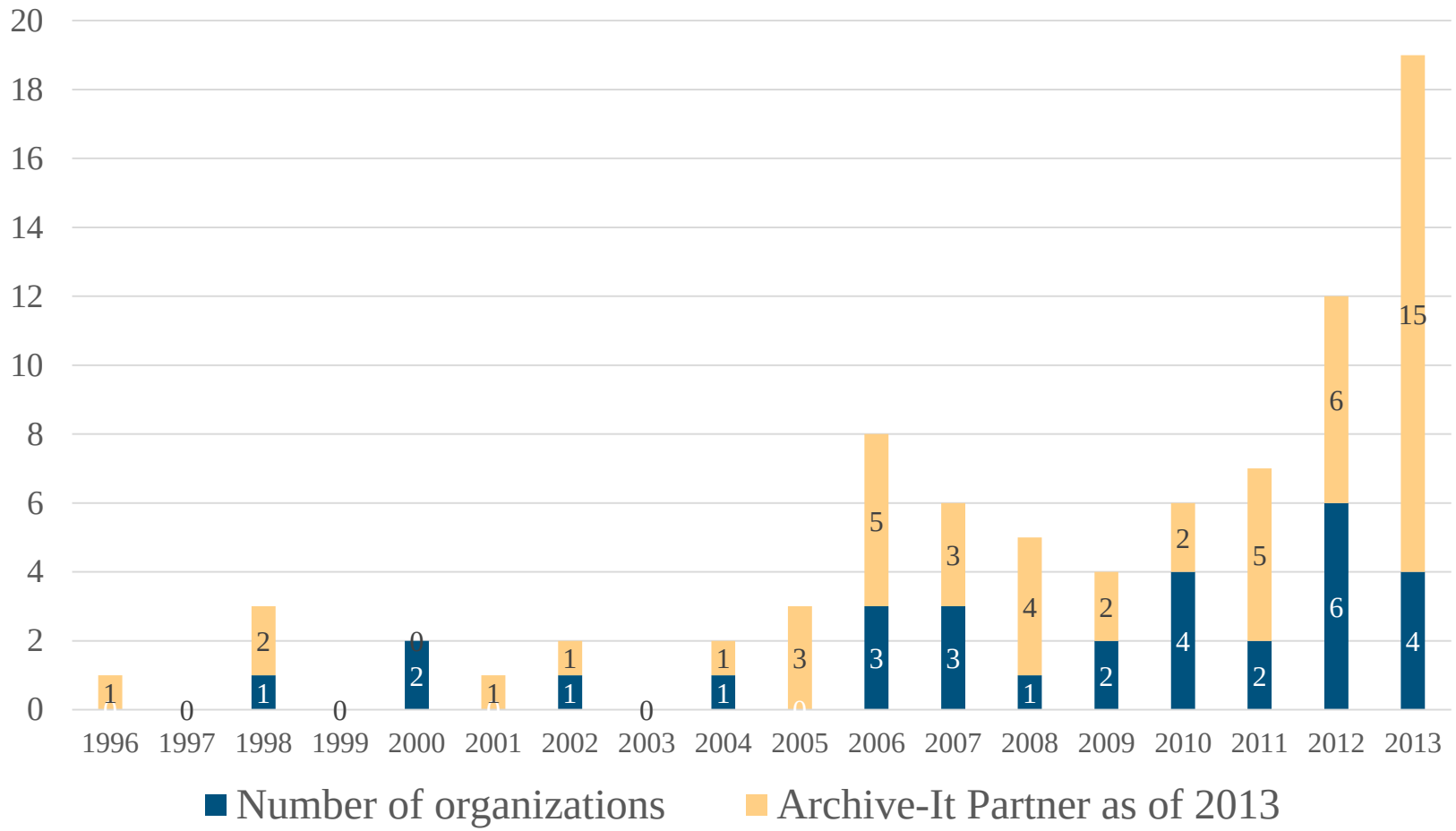
Andrew Jackson: "Ten years of the UK Web Archive"

a centralized enterprise



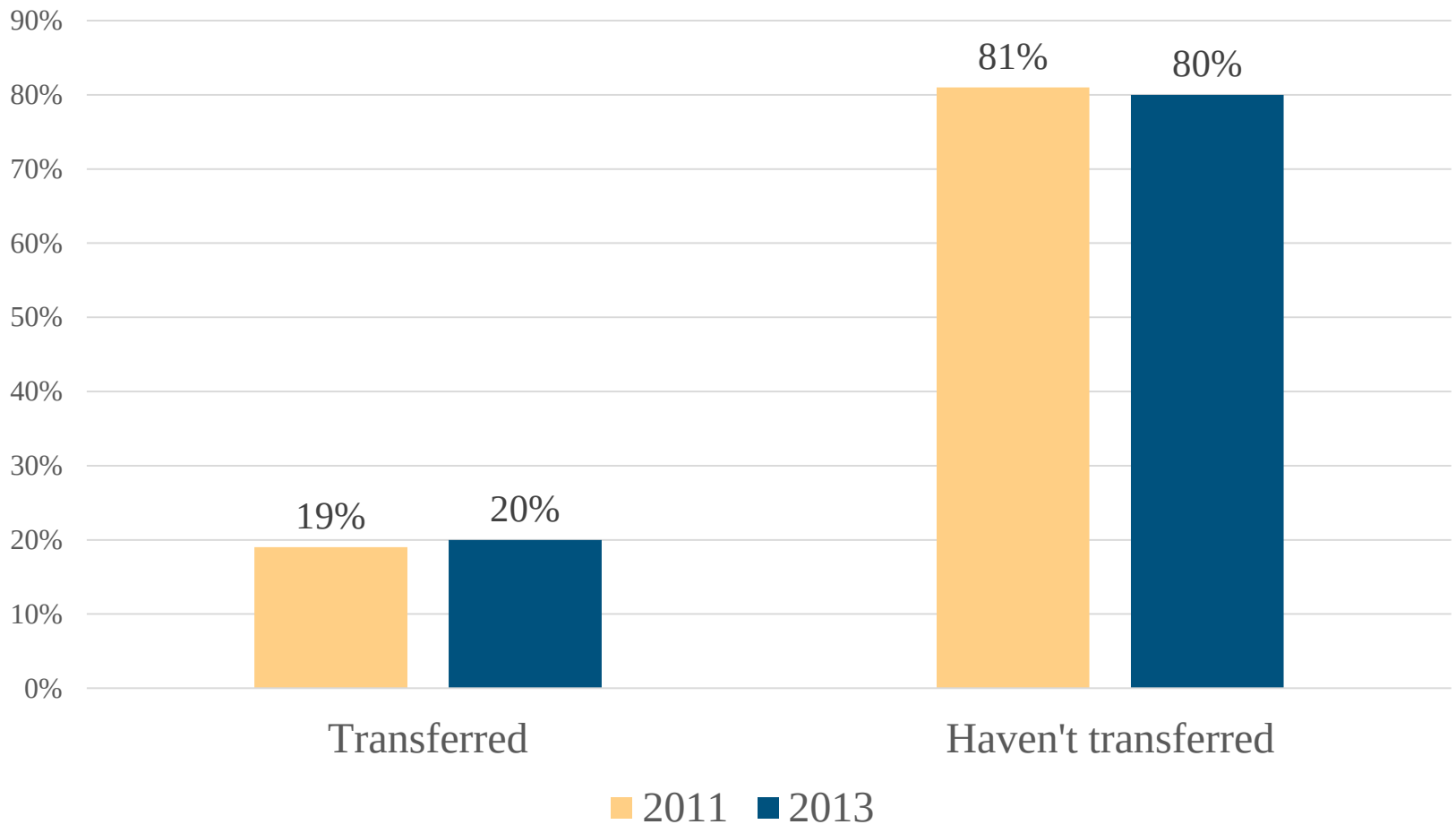
NDSA: "Web Archiving in the U.S.: A 2013 Survey"

a centralized enterprise



NDSA: "Web Archiving in the U.S.: A 2013 Survey"

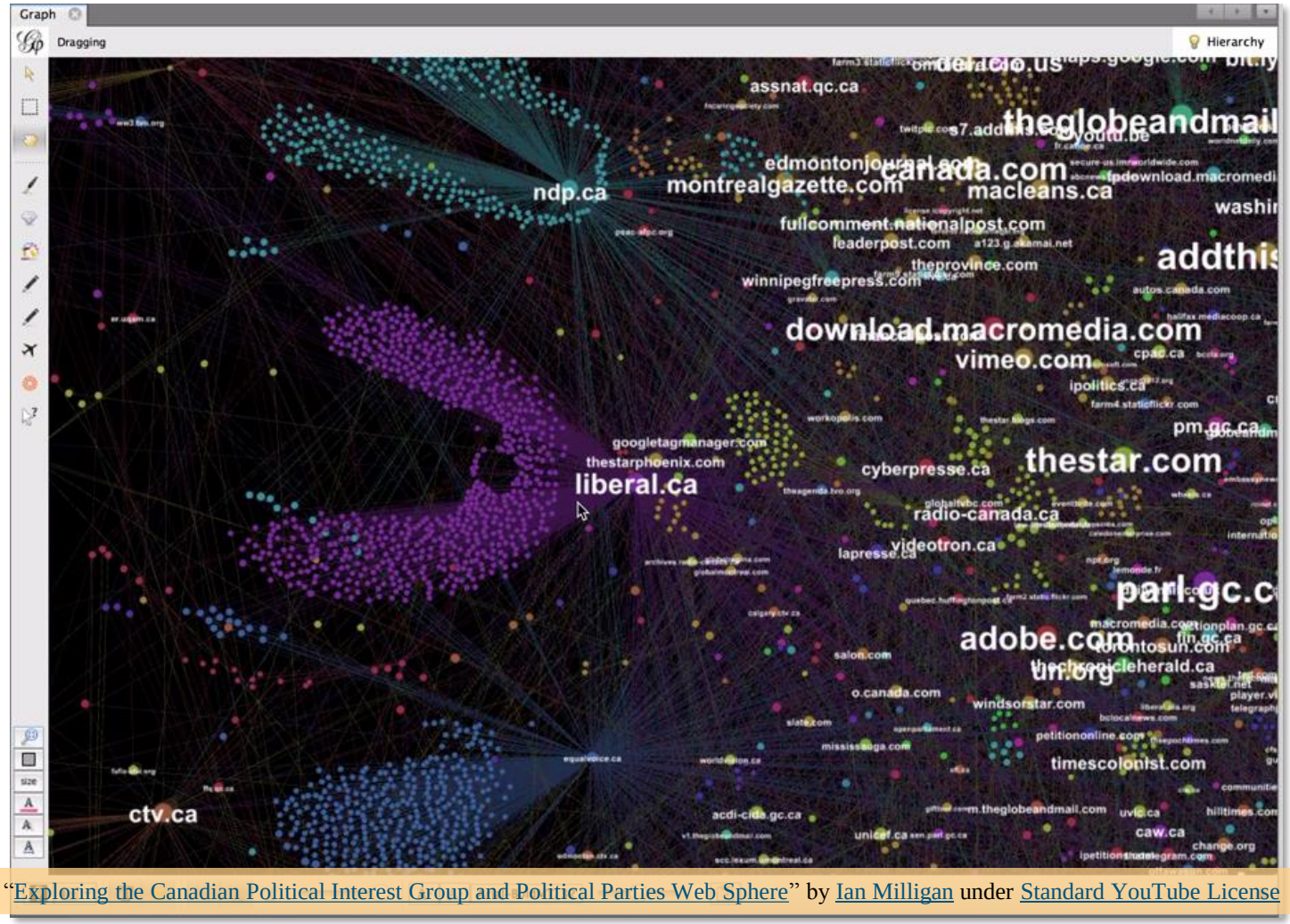
minimal local preservation



NDSA: "Web Archiving in the U.S.: A 2013 Survey"

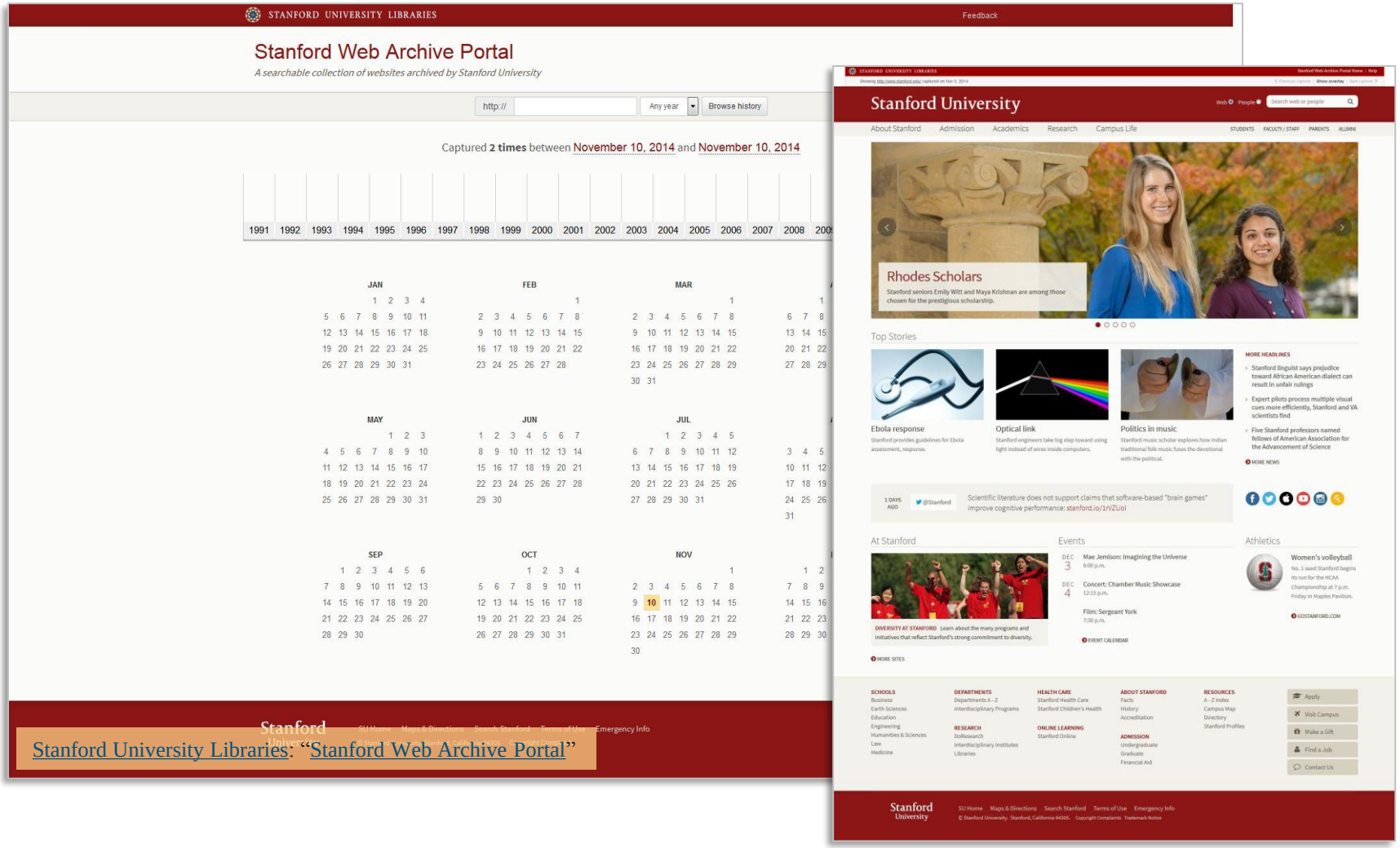
evolving web

opportunities for research



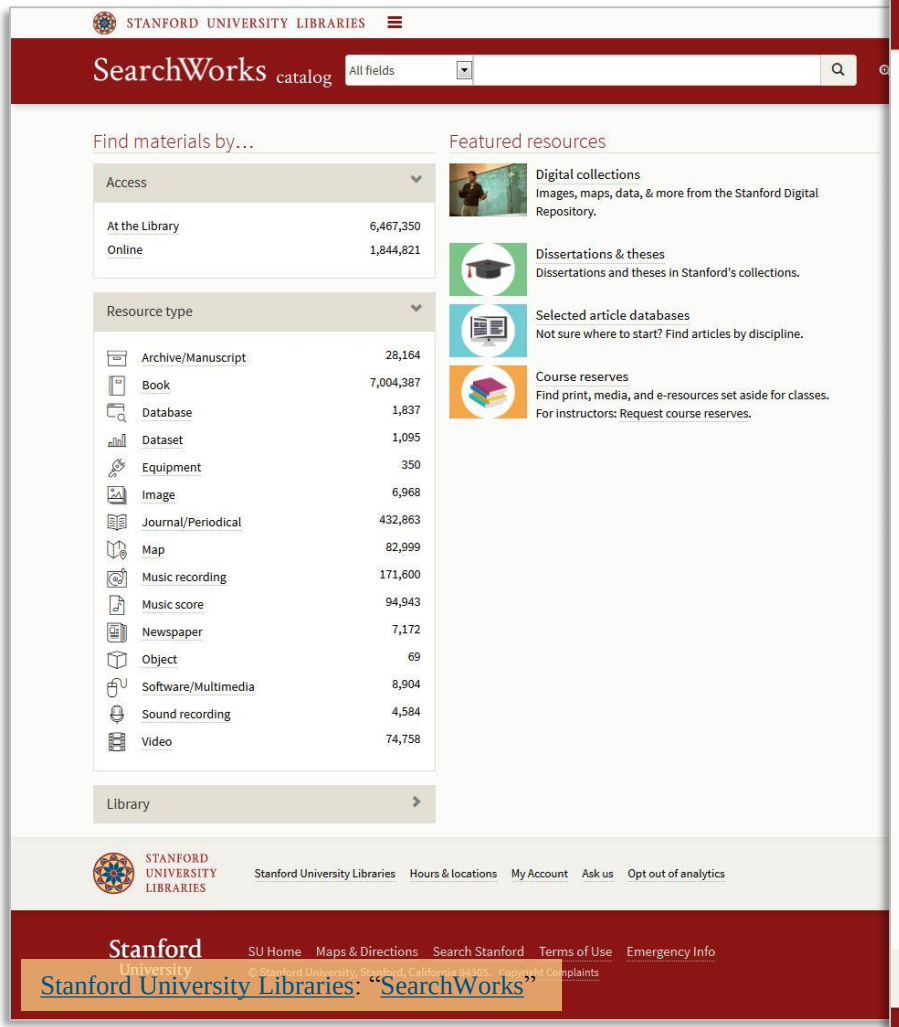
WHAT ARE WE DOING?

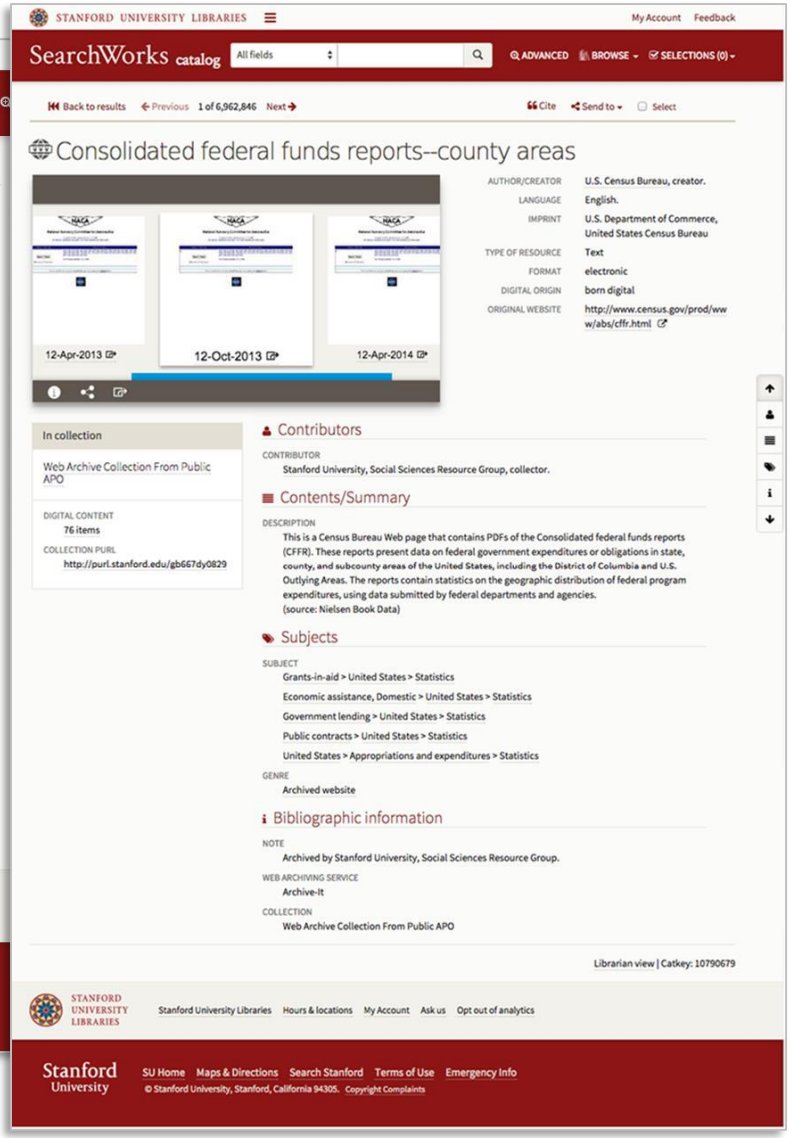
Stanford Web Archive Portal



Stanford University Libraries: "Stanford Web Archive Portal"

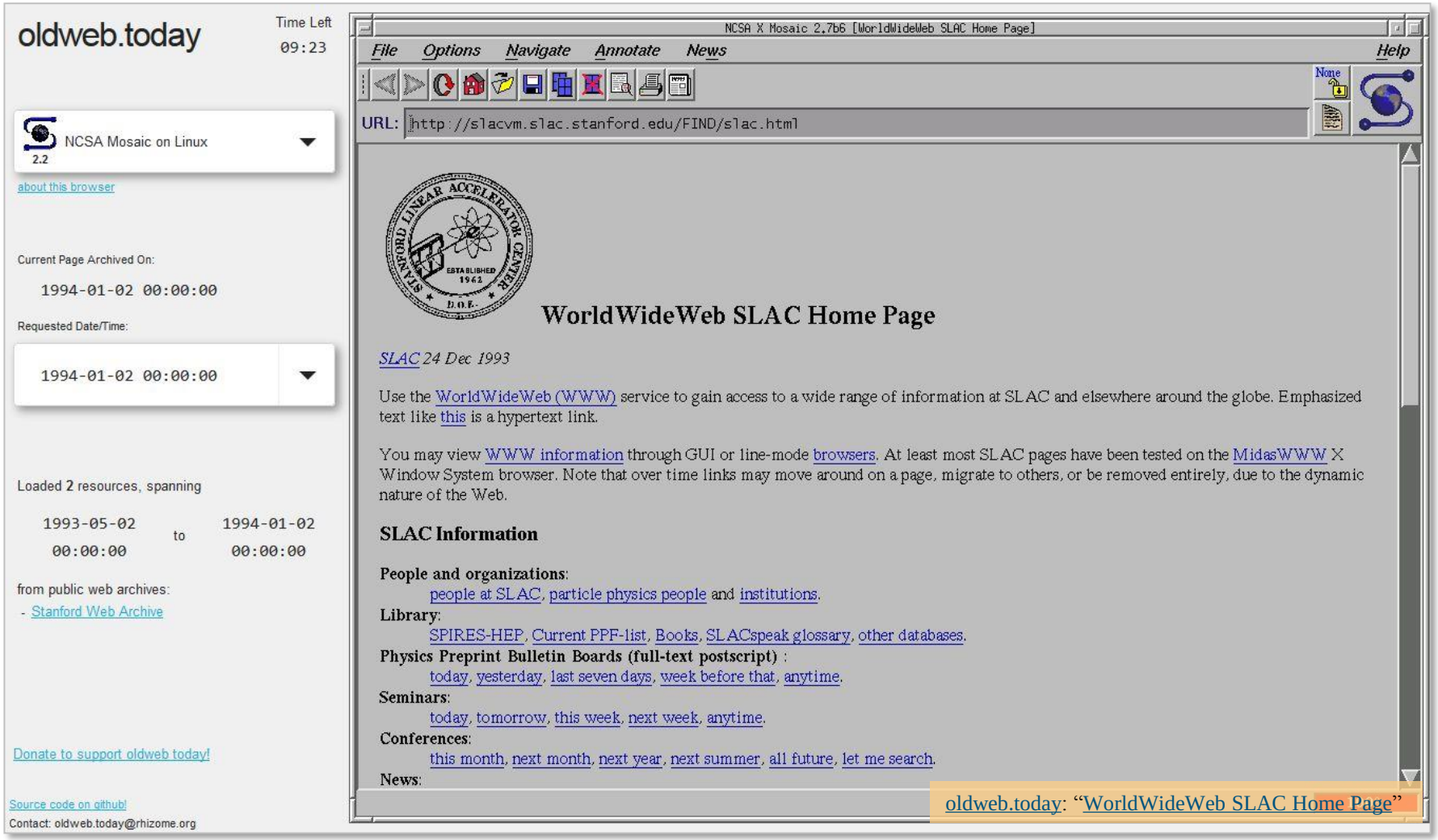
SearchWorks (online catalog)





Stanford University Libraries: "SearchWorks"

web archaeology (SLAC)



oldweb.today: "WorldWideWeb SLAC Home Page"

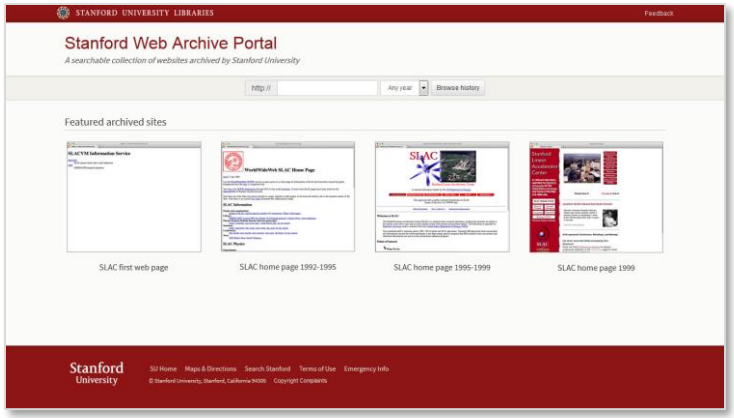
building + integrating infrastructure

ARCHIVE-IT

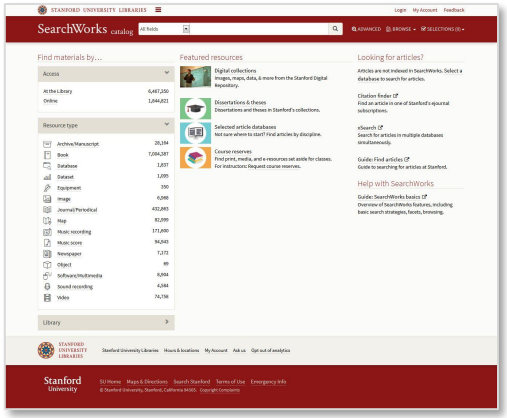
capture

SDR

preservation



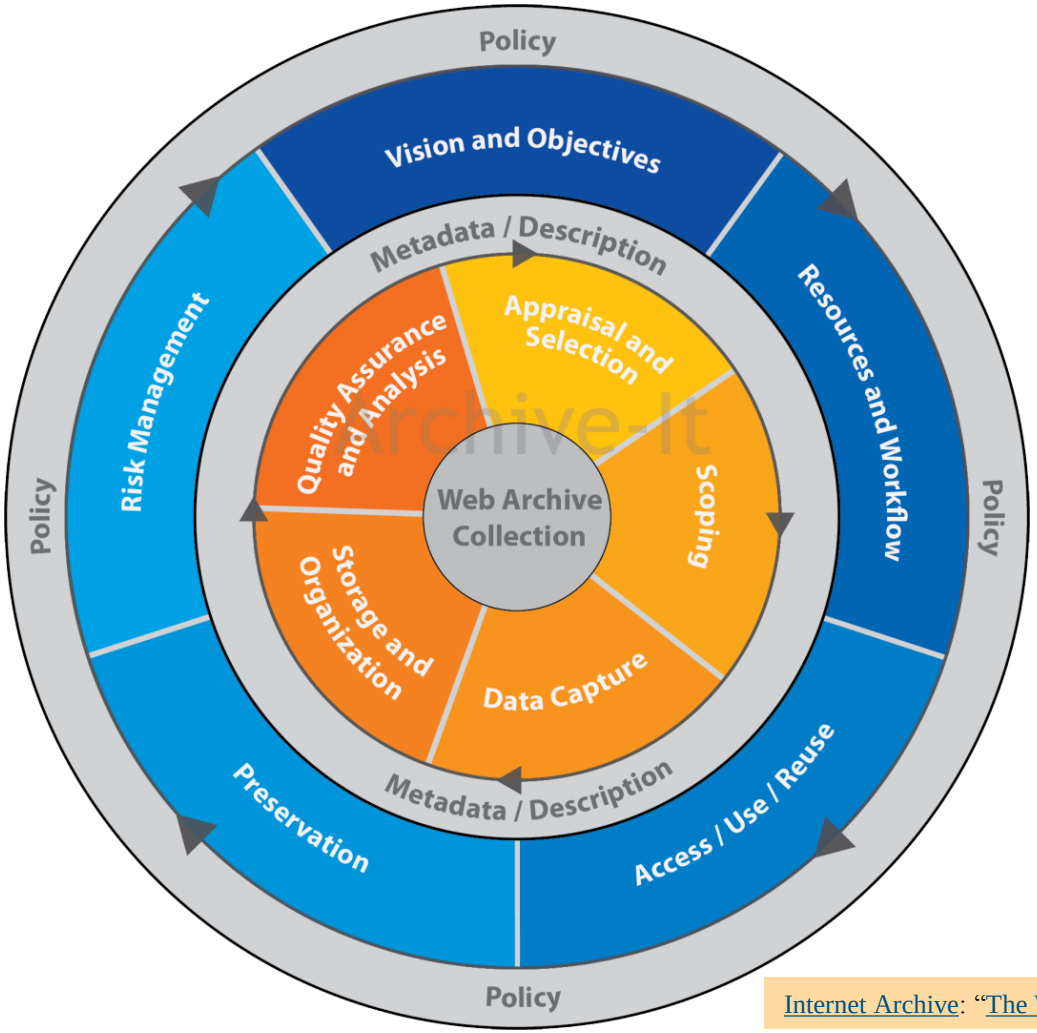
access



discovery

APIS + COMMUNITY DEVELOPMENT

web archiving lifecycle

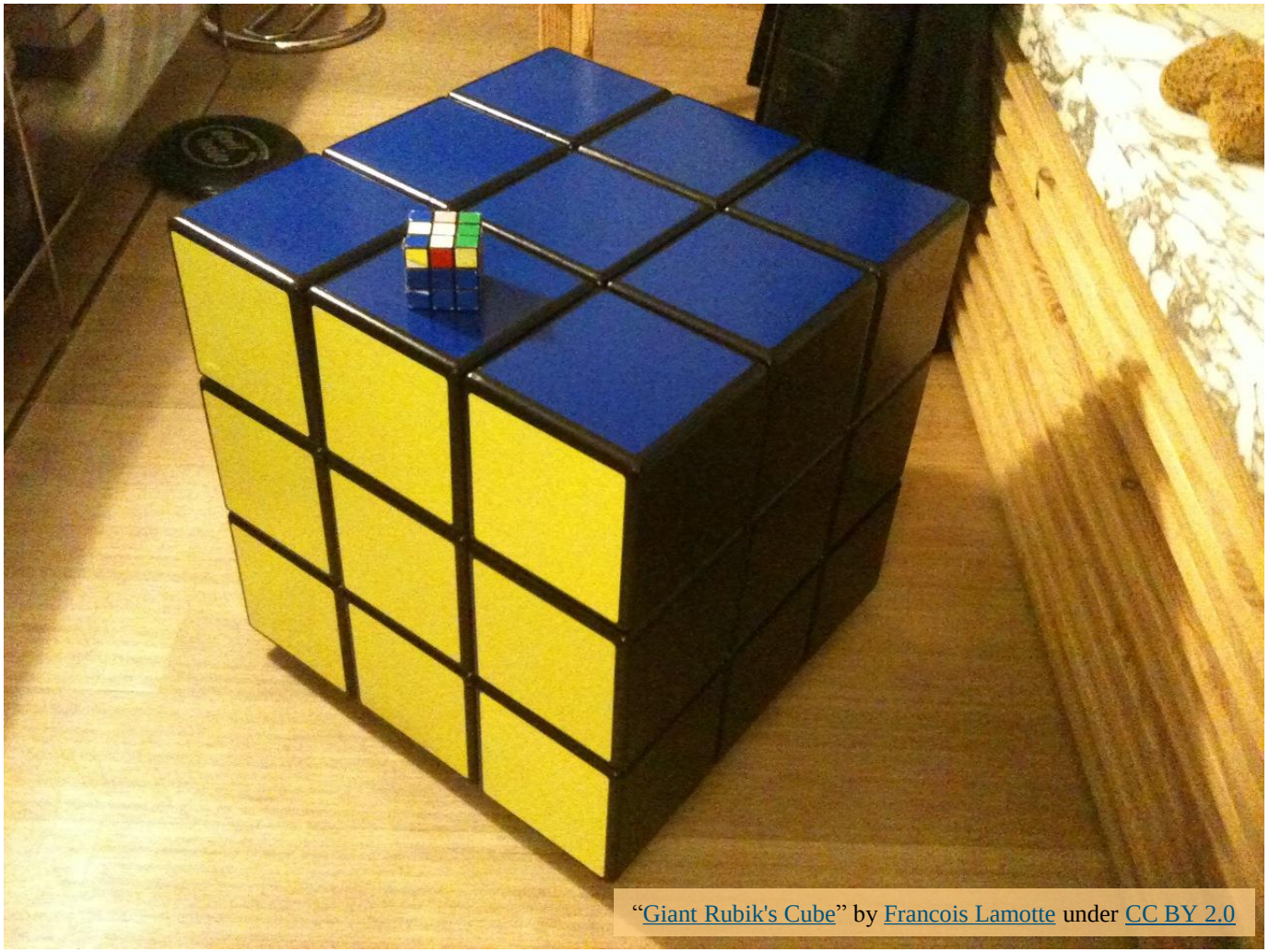


Internet Archive: "The Web Archiving Life Cycle Model"

functional overlap

	Appraisal and Selection	Scoping	Data Capture	Storage and Organization	QA and Analysis	Metadata / Description	Access / Use / Reuse	Preservation	Risk Management
ACT									
Archive-It									
AtN									
BCWeb									
CDL WAS									
DigiBoard									
Islandora WARC Solution Pack									
Netarchive Suite									
PageFreezer									
UNT Nomination Tool									
WCT									

smaller, modular components



community seed

National Digital Platform Projects funded in August 2015

Systems Interoperability and Collaborative Development for Web Archiving (LG-71-15-0174-15): The Internet Archive, working with partner organizations University of North Texas, Rutgers University, and Stanford University Library will undertake a two-year research project to explore techniques that can expand national web archiving capacity in several areas.

INTERNET ARCHIVE

INSTITUTE of Museum and Library SERVICES

RUTGERS UNIVERSITY

UNT

API candidates

- capture tool/proxy interconnect
- capture tool management
- data import/export
- query + extraction
- integrity audit + repair
- descriptive metadata
- logs + analytics
- renderings/derivative formats
- federated data delivery
- federated replay
- federated full-text search

let's combine forces