



Center *for* Research Libraries  
GLOBAL RESOURCES NETWORK

# CLOCKSS, LOCKSS and the "Long Tail," Preserving at- risk Publishers

Nicholas Taylor, Victoria Reich  
Stanford University Libraries

# introductions



# strong alignment



- LOCKSS Program builds technical solutions, enabling communities to collect, preserve, and ensure access to uncommon/unique collections - the “long tail”
- CRL community builds unique/uncommon collections and ensures its long-term integrity and accessibility for CRL affiliated scholars

# how many titles?



- count journals meeting industry expectations
  - ISSN, peer review
- estimates differ
  - 60,000 STM - Science-Metrix
    - osi-list 9/12/16, 4:28 AM <http://www.science-metrix.com/en>
  - 35,000 STM - Mark Ware report
    - [http://www.stm-assoc.org/2015\\_02\\_20\\_STM\\_Report\\_2015.pdf](http://www.stm-assoc.org/2015_02_20_STM_Report_2015.pdf)
- doesn't count
  - developing country titles
  - humanities/cultural titles
  - many government documents

# the long tail defined



- Why the definition matters
  - scopes the problem
  - needed to appropriately focus resources
- *“publishers with less than 10 journals”*
  - generous definition
  - deflects tool and process optimization
- alternative definition would include
  - subject to government censorship
  - disseminated by immature organizations
  - other criteria?

# overcoming the challenges



money

....and determination



# challenge: social



- post cancellation access (pca) motivates preservation
  - preserve access to most expensive content
  - open access titles have no pca motivation
- identifying titles
  - shift away from subject specialists
  - collection development skills
    - needed: basic bibliographic knowledge and curiosity

# challenge: technical



- Ingest is single biggest cost
  - 3000 titles/publisher = 1 title/publisher
  - big publishers have uniform, automated processes
  - small publishers have idiosyncratic processes
- publishers need sophistication to
  - push content to preservation service
- publishers need fewer resources to
  - put online permission text statement (or CC license)



# challenge: business



- publishers
  - contracts
  - fees
- Librarians
  - need rewards and incentives to take action

# solution



- drive down costs
  - further alignments with web archiving technology
- share remaining costs as widely as possible
  - partner with cognizant communities

# philosophy



Preservation is an  
*active*  
community effort

LOTS OF COPIES  
KEEP STUFF SAFE

LOTS OF COMMUNITIES  
KEEP STUFF SAFE



# the LOCKSS Program



**Communities** use  
LOCKSS open  
source software to  
preserve **their**  
scholarly and cultural  
record



# TRAC/ISO16363 audit



- 1<sup>st</sup> ever **perfect score** for Technologies, Technical Infrastructure, Security
- documentation linked from dshr.blog.org
  - Introduction <http://blog.dshr.org/2014/07/trac-certification-of-clockss-archive.html>
  - TRAC Audit: Process - <http://blog.dshr.org/2014/08/trac-audit-process.html>
  - TRAC Audit: Lessons - <http://blog.dshr.org/2014/08/trac-audit-lessons.html>
  - TRAC Audit: Do-It-Yourself Demos - <http://blog.dshr.org/2014/08/trac-audit-do-it-yourself-demos.html>

# many LOCKSS networks



- 1000+ of publishers
  - Subscription, open access, etc.
- Ingest techniques
  - OAI/PMH, web crawling, file transfer, API, etc.
- Preserved content types
  - Journals, books, databases, government documents, thesis and dissertations, image collections...
- Each with an organization, business model
- Each with an appropriate access policy

[www.clockss.org](http://www.clockss.org)



# CLOCKSS collection development



*ensure the long-term survival and prospect of access to web-based publications\* of value for scholarship worldwide*

- any and all languages
- full-text and all supplementary materials

\*from reputable and ethical publishers

- scholarly publication norms (e.g. issns/isbns/DOIs)

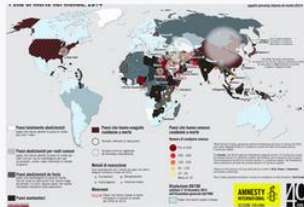


# CLOCKSS long tail examples



Cartografare  
il presente

HOME CHI SIAMO CARTOTECA DOSSIER ATTUALITÀ DAL MONDO ATLANTE



03/05/2015

## La pena di morte nel 2014

Nieves López Izquierdo

Il rapporto di Amnesty International sulla pena di morte nel mondo nel 2014 rileva due fenomeni in potenziale contrasto. Da un lato, i paesi che hanno eseguito condanne a morte sono stati 22 - lo stesso numero del 2013. Dall'altro, le condanne a morte sono aumentate drasticamente: almeno 2466 a causa soprattutto di Egitto e Nigeria.

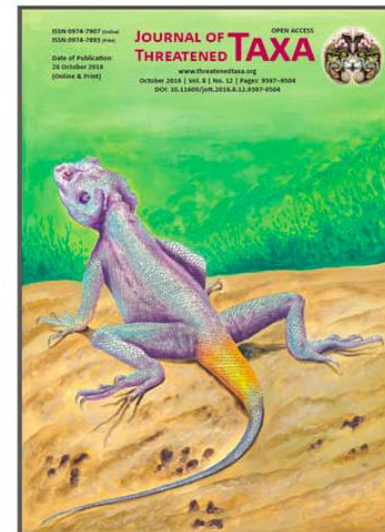
[Leggi Tutto...](#)

Big enough to

- sign a contract
- pay a fee



The International Journal on Conservation & Taxonomy  
ISSN: 0974-7907 (online); 0974-7893 (print)



# Brazil's Cariniana



IBICT - Instituto Brasileiro de Informação em Ciência e Tecnologia

**Rede Cariniana**

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO

IBICT

Twitter YouTube Facebook

Agenda Contatos Eventos Mapa da Rede Notícias Perguntas frequentes

ALIANÇA LOCKSS

**SOBRE**

Início

IBICT

Preservação Digital

Aliança LOCKSS

**Parcerias da Rede**

Parcerias Integrals

Parcerias Institucionais

Logos por Participantes

**SERVIÇOS**

**Rede de Pesquisa ORIADE**

Fóruns

**Preservação de documentos**

Periódicos Eletrônicos

**Cursos técnicos**

Curadoria Digital

Preservação Digital

**PRODUTOS**

Dicionário de Preservação Digital

## Rede Cariniana e LOCKSS



Aliança formada em prol da preservação digital do patrimônio científico e cultural



**Programa LOCKSS**

Ferramenta de preservação digital criada pela equipe da biblioteca da Universidade de Stanford que utiliza uma arquitetura distribuída



**Rede Privada LOCKSS**

Rede Privada LOCKSS com parcerias de Instituições de todo Brasil



**Modelos de Preservação Digital**

Distribuir ou não distribuir, eis a questão


# PKP Private LOCKSS Network





# U.S. Government Documents





LOTS OF COPIES  
KEEP STUFF SAFE

Search

About LOCKSSThe CommunityNews & MediaTechnical SupportHow to JoinContact Us

LOCKSS > The Community > Global & Private LOCKSS Networks > Digital Federal Depository Library Program

## Digital Federal Depository Library Program

The Digital Federal Depository Library Program, or the "USDocs" private LOCKSS network replicates key aspects of the United States Federal Depository System. The content is held in geographically distributed sites and replicated many times. Citizens have oversight and responsibility for the long-term care and maintenance of the content. All these characteristics mean the content will be preserved so that any alteration of the content (either deliberate or accidental) will be detected and repaired. For all documents, preservation in a tamper-evident environment is important, but for government documents, this is essential.

Our current focus is harvesting and preserving collections hosted on the The U.S. Government Printing Office's (GPO) Federal Digital System (FDsys). See also, Digital Federal Depository FAQ

### Participating Institutions

- Alaska State Library
- Amherst College
- Arizona State Library, Archives & Public Records [FDLP regional library]
- California State University, Fresno
- George Washington University
- Georgetown University
- Georgia Institute of Technology
- Government Printing Office
- Idaho State University
- Indiana State Library [FDLP regional library]
- Library of Congress
- Louisiana State University [FDLP regional library]
- Louisiana Tech University [FDLP regional library]
- Michigan State University
- North Carolina State University
- Northwestern University
- Princeton University
- Rice University
- San Diego State University
- Simon Fraser University
- Stanford University
- University of Alabama [FDLP regional library]

4/2013  
Three new case studies, U.K. LOCKSS Program, Ensuring preservation and access.

Private LOCKSS Networks. Communities of all sizes are using Private LOCKSS Networks to preserve their own cultural heritage materials. Join one, or build your own!



James Jacobs, Stanford  
jrjacobs at stanford dot edu

# Canadian Government Information



## CGI network

From Plnwiki

### Canadian Government Information (CGI) LOCKSS network

Contents [\[hide\]](#)

- [1 Participating Institutions](#)
- [2 Governance](#)
- [3 Collection and Access Policy](#)
- [4 Steering Committee](#)
- [5 Related Resources](#)

## Participating Institutions

- Dalhousie University
- McGill University
- Scholar's Portal
- Simon Fraser University
- Stanford University
- University of Alberta
- University of British Columbia
- University of Calgary
- University of Saskatchewan
- University of Toronto
- University of Victoria

[nt](#)

[\[1\]](#) **Press Release, September 2013**

## Governance

### Canadian Government Information Digital Preservation Network

#### Governance Policy

#### 1.0 Introduction and Mission

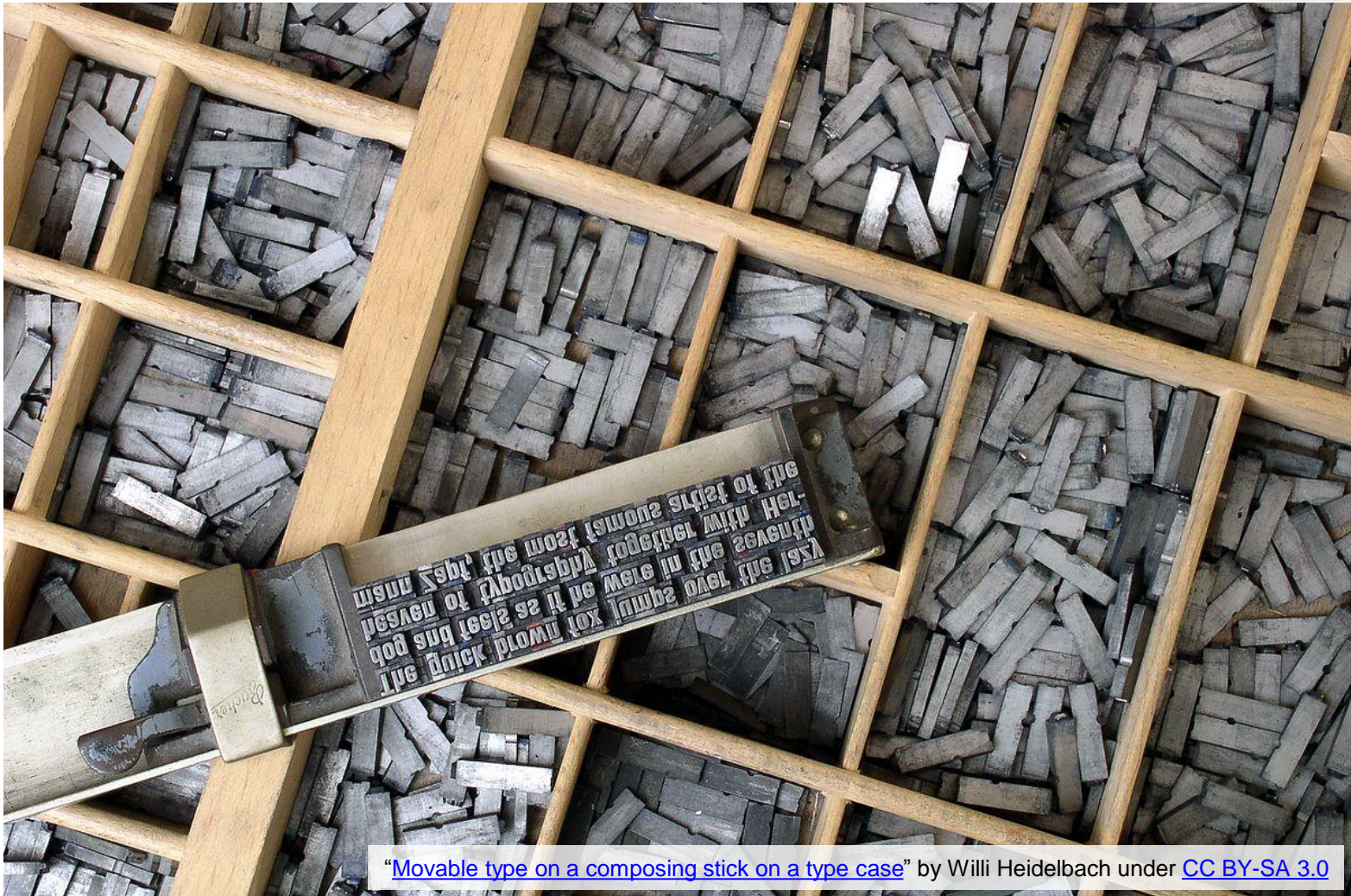
##### 1.1 The Canadian Government Information Digital Preservation Network

- The Canadian Government Information Digital Preservation Network (CGI DPN) is a project initiated in October 2012 by library staff at eleven member institutions: University of Alberta, Simon Fraser University, University of British Columbia, University of Calgary, University of Saskatchewan, University of Victoria, McGill University, Dalhousie University, Scholars Portal, University of Toronto, Stanford University. These institutions are the charter members of the CGI DPN Steering Committee.

##### 1.2 Mission and Purpose



# web publishing



"Movable type on a composing stick on a type case" by Willi Heidelbach under [CC BY-SA 3.0](https://creativecommons.org/licenses/by-sa/3.0/)

# mixed impact for preservation



## benefits

- more broadly accessible
- archived content can remain accessible
- archived content still “belongs to” the Web
- amenable to collective action

## challenges

- volume
- discovery
- diversity of platforms
- finite window for archiving
- concentrated costs for archiving



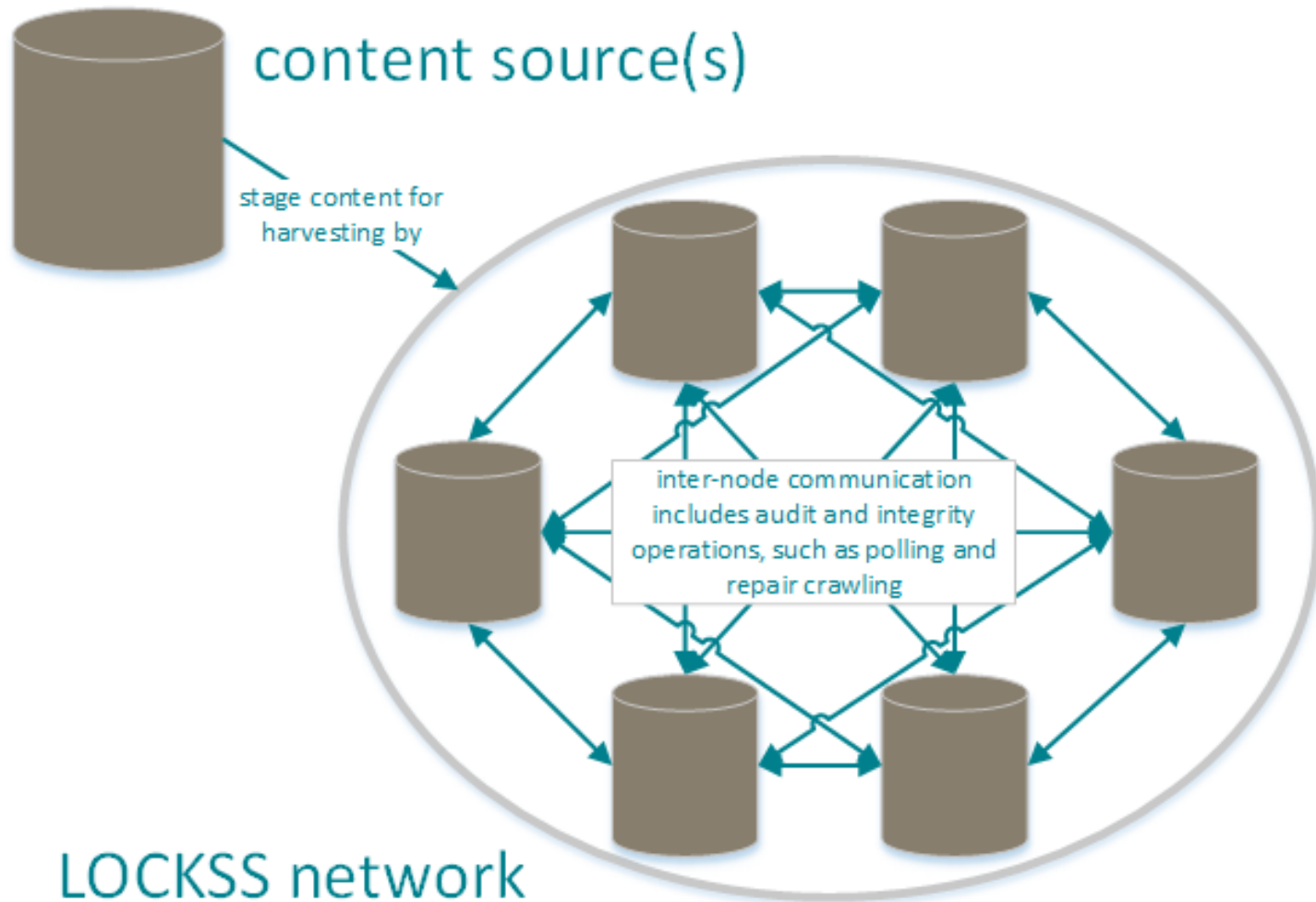
# what to collect?



“The Seeker” by [C MB 166](#) under [CC BY-ND 2.0](#)



# LOCKSS architecture



# changing web content



2001

2005

2016

Case report  
Organisational development in general practice: lessons from practice and professional development plans (PPDPs)

Glyn Elwyn1 email and Paul Hocking2 email

1  
Department of General Practice University of Wales College of Medicine Canolfan Iechyd Llanedeyrn Health Centre CF23 9PN

2  
Primary Care Development Research Group NHS Staff College Wales Hensol Castle Pontyclun Mid Glamorgan CF72 8YS

BMC Family Practice 2000, 12:doi:10.1186/1471-2296-1-2  
Published: 19 December 2000  
Abstract  
Background

Improving the quality and effectiveness of clinical practice is becoming a key task within all health services. Primary medical care, as organised in the UK is composed of clinicians who work in independent partnerships (general practices) that collaborate with other health care professionals. Although many practices have successfully introduced innovations, there are no organisational development structures in place that support the evolution of primary medical care towards integrated care processes. Providing incentives for attendance at passive educational events and promoting 'teamwork' without first identifying organisational priorities are interventions that have proved to be ineffective at changing clinical processes. A practice and professional development plan feasibility study was evaluated in Wales and provided the experiential basis for a summary of the lessons learnt on how best to guide organisational development systems for primary medical care.

Results

Practice and professional development plans are hybrids produced by the combination of ideas from management (the applied behavioural science of organisational development) and education (self-directed adult learning theories) and, in conceptual terms, address the lack of effectiveness of passive educational strategies by making interventions relevant to identified system wide needs. In the intervention, each practice participated in a series of multidisciplinary workshops (minimum 4) where the process outcome was the production of a practice development plan and a set of personal portfolios, and the final outcome was a realised organisational change.

It was apparent during the project that organisational admission to a process of developmental planning needed to be a stepwise process, where initial interest can lead to a fuller understanding, which subsequently develops into motivation and ownership, sufficient to complete the exercise. The advantages of introducing expert external facilitation were clear: evaluations of internal group processes were possible, strategic issues could be raised and explored and financial probity ensured. These areas are much more difficult to examine when only internal stakeholders are engaged in a planning process.

Conclusions

It is not possible to introduce practice and professional development plans (organisational development and organisational learning projects) in a publicly funded health care system without first addressing existing educational and management structures. Existing systems are based on educational credits for attendance and emerging accountability frameworks (criteria checklists) for clinical governance. Moving to systems that are less summative and more formative, and based on the philosophies of continual quality improvement, require changes to be made in the relevant support systems in order to achieve policy proposals.

# aligning with web archiving



## Web ARChive file format



## compatible technologies

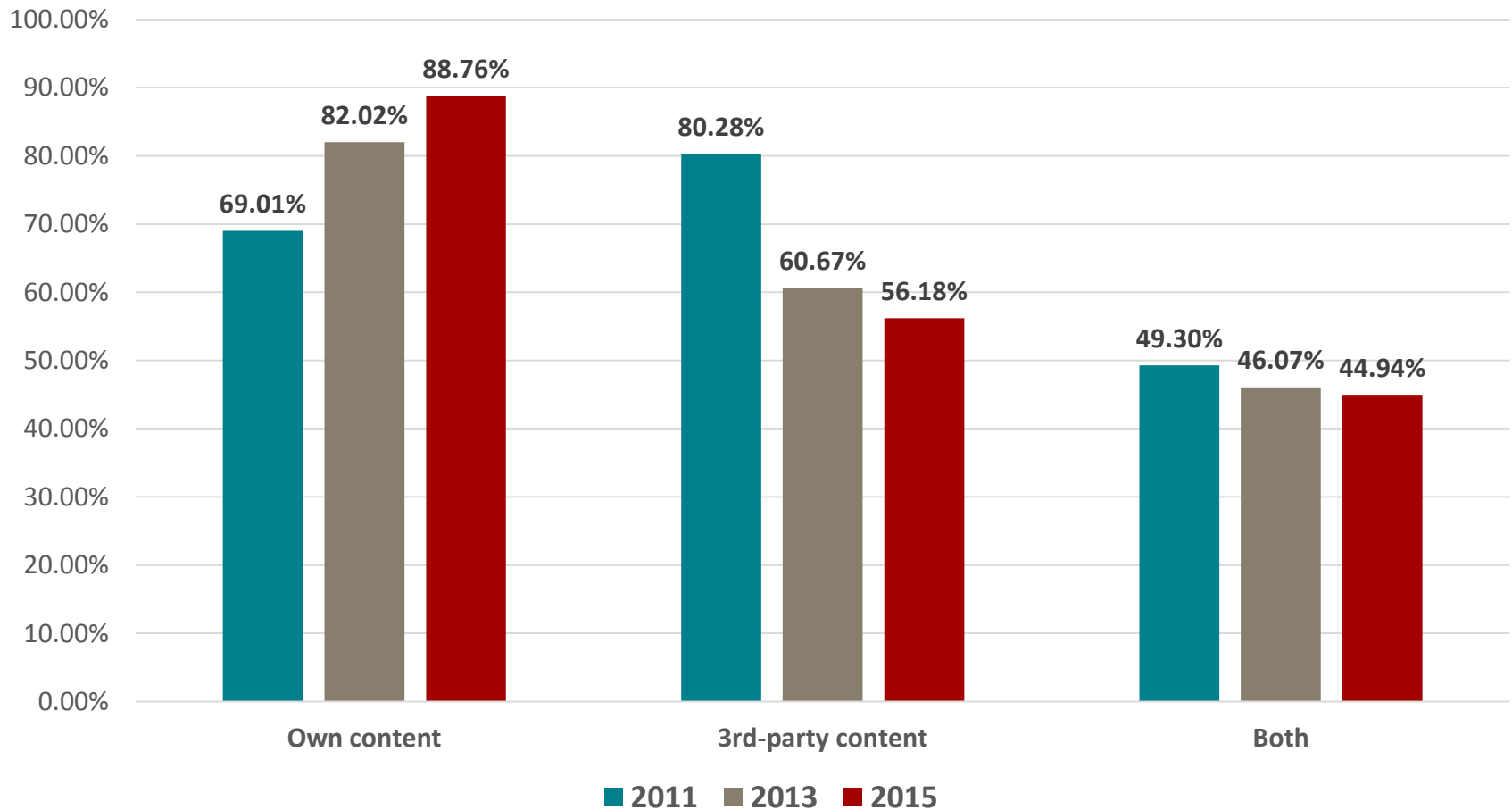
- Heritrix
- OpenWayback
- WarcBase
- WASAPI Data Transfer APIs
- Web Archiving Proxy

# reducing ingest costs



- prepare to evolve w/ the Web
  - web services architecture as flexible foundation
  - collaborate on specifying web archiving APIs
- don't build it all ourselves
  - leverage web-scale open-source software
  - enlist + fortify related technical communities
- de-silo components for external re-use
  - metadata extraction
  - archive access via DOI + OpenURL
  - polling + repair protocol

# collective investment in web archiving



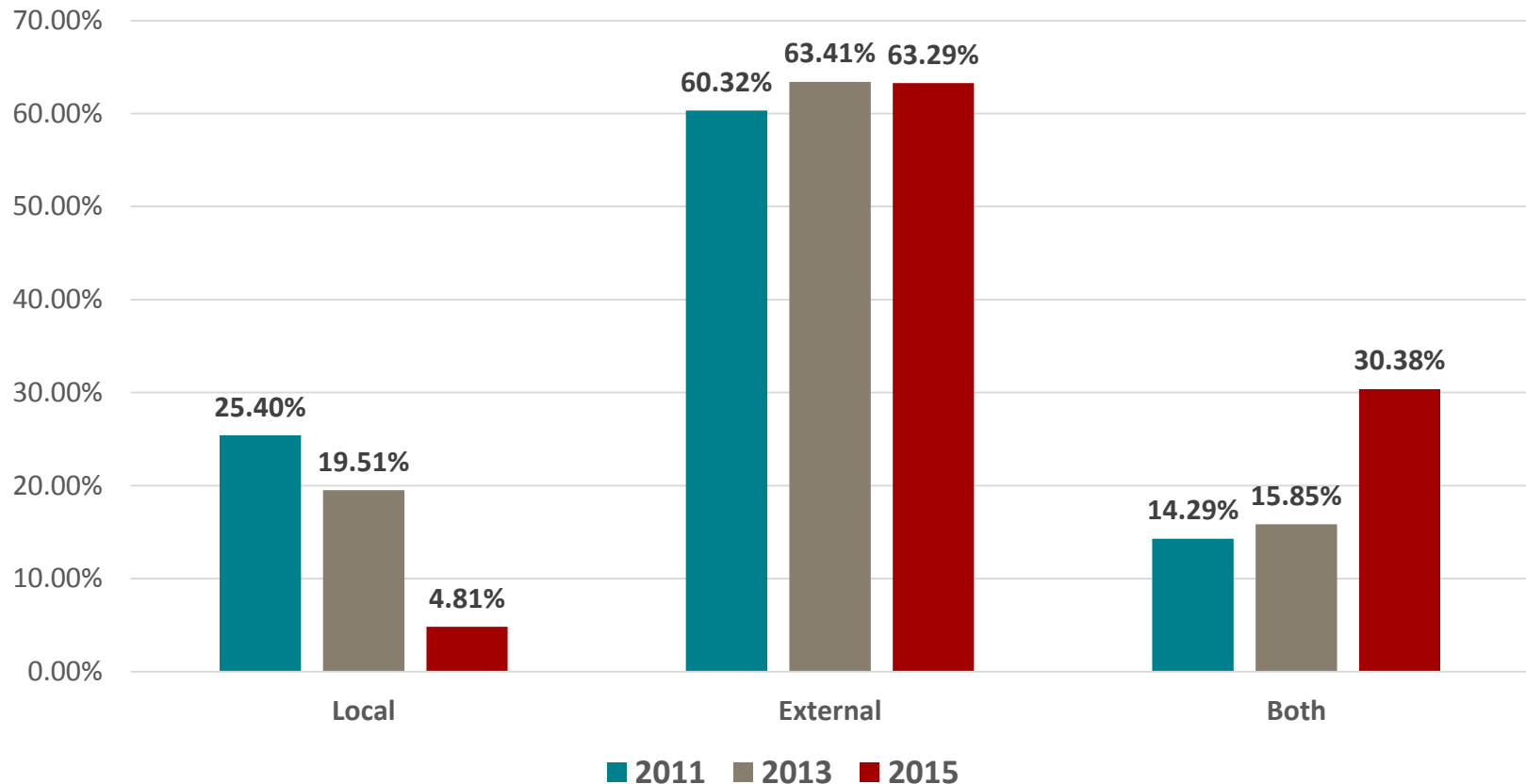
[NDSA](#): "2015 NDSA Web Archiving Survey"

# coordinated web content collecting



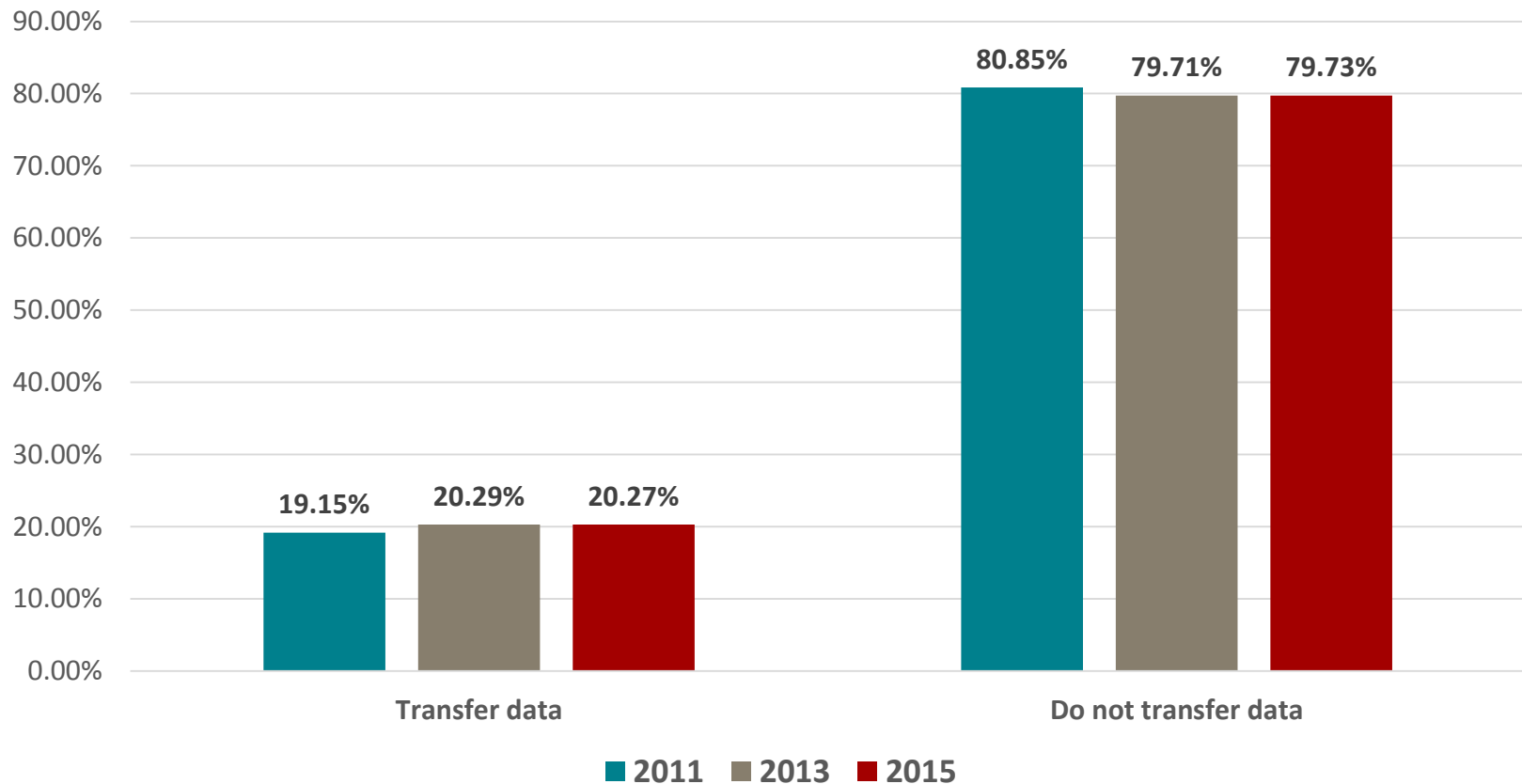
- better match collective resources + collection development expertise to scale of challenge
- examples of initiatives
  - CDL, Harvard, + UCLA (w/ IMLS funding) [Cobweb project](#)
  - IIPC [Content Development Working Group](#)
  - Ivy+ Art and Architecture Group Collaborative Architecture, Urbanism, and Sustainability Web Archive ([CAUSEWAY](#))
  - Ivy+ Collaborative Web Content Collecting Initiative

# reliance on service provider



[NDSA](#): "2015 NDSA Web Archiving Survey"

# flat data transfer trend



[NDSA](#): "2015 NDSA Web Archiving Survey"



# distributed preservation



- ensure diversity of geographic, organizational, technical contexts for preserving archived web
- examples of initiatives
  - Ivy+ Preservation Network
  - ArchiveTeam [IA.BAK](#) initiative
  - Internet Archive, Stanford, Rutgers, UNT project to develop web archive data transfer APIs ([WASAPI](#))

# questions for the group



- which **subject areas that you support** have long-tail content?
- provided the technical means, **would you be interested in collecting** that long-tail content?
- who else – collaborators, stakeholders – **shares your interest** in collecting that long-tail content?
- what are the **opportunities for working together** to preserve long-tail content?

# Thank You

