Preserving News and Journalism:

Fundamentals of Born Digital News Archiving

http://bit.ly/1kIZ4f2

Digital Preservation 2014

Introductions

Panel

Anne Wootton
Pop Up Archive
@annewooton

Leslie Johnston
Director of Digital Preservation - NARA
@llohnston

Edward McCain
Reynolds Journalism Institute, University of Missouri
@e_mccain

Aurelia Moser
Knight Mozilla Open News Fellow
@auremoser

Intro to Newsroom Preservation

Terms:

- What is ‘born digital’ in a newsroom?
- What are ‘news apps’?

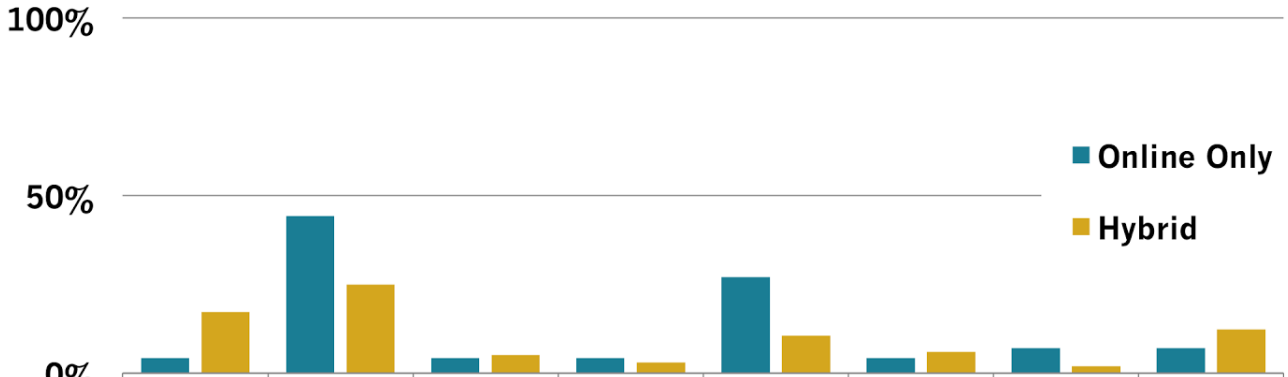
Survey Results

RJI phone survey of 476 news organizations:

- 406 were “Hybrid” enterprises
- 70 were “Online Only” publishers

Survey Results

Born-digital news text formats



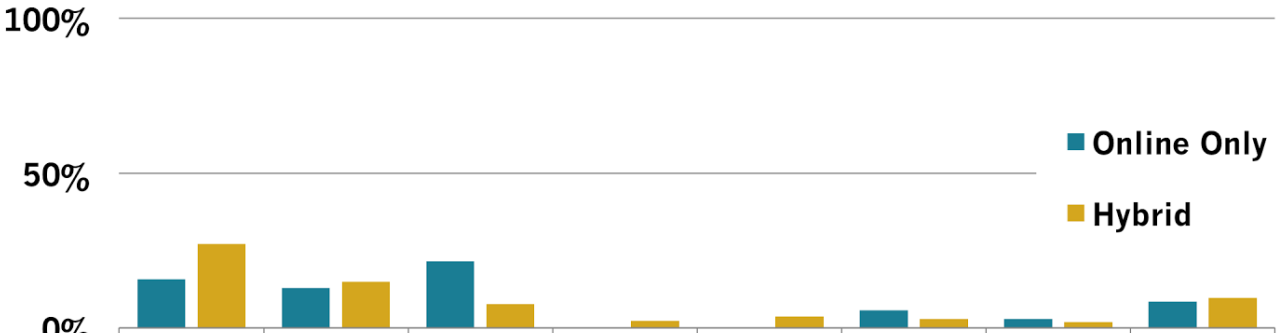
	XML	HTML	ASCII	UTF-8	WORD	PDF	WEB	OTHER
	Q1aa: In what format is the BD text content produced?							
Online Only	4%	44%	4%	4%	27%	4%	7%	7%
Hybrid	17%	25%	5%	3%	11%	6%	2%	12%

jdna journalism digital news archive

rji DONALD W. REYNOLDS JOURNALISM INSTITUTE
Missouri School of Journalism University of Missouri

Survey Results

Born-digital news video formats



Q1cc: In what format is the BD video content produced?

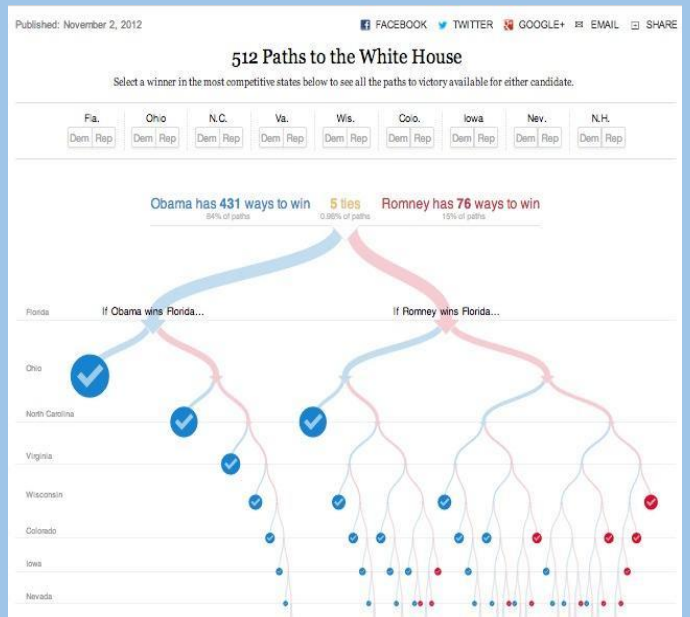
	MP4	MOV	YOU-TUBE	BRIGHTCOVE	MPEG	FLASH	AVI	OTHER
Online Only	16%	13%	21%	0%	0%	6%	3%	9%
Hybrid	27%	15%	8%	2%	4%	3%	2%	10%

jdna journalism digital news archive

DONALD W. REYNOLDS JOURNALISM INSTITUTE
Missouri School of Journalism University of Missouri

News Apps

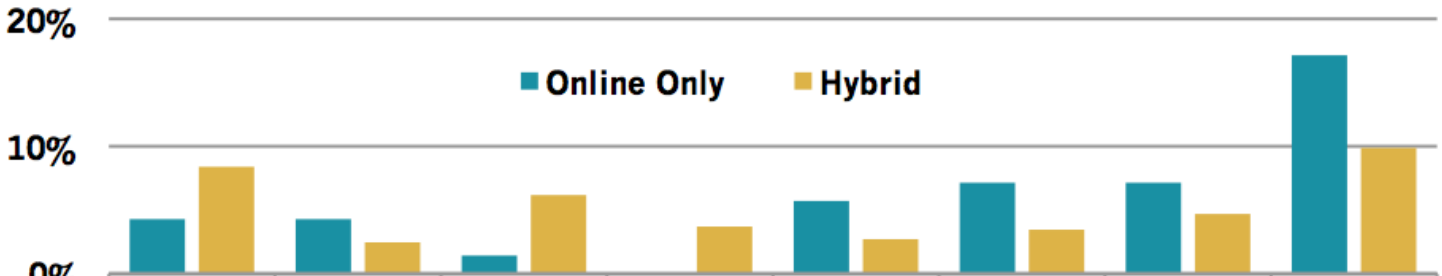
ProPublica | Dollars for Docs | 3/3/14
http://bit.ly/1iJD2cq



NYTimes | 512 Paths to White House | 11/2/12
http://nyti.ms/OlWyxL

Survey Results

Born-digital news interactive formats



Q1dd: If so, in what format is the BD Interactive content produced?

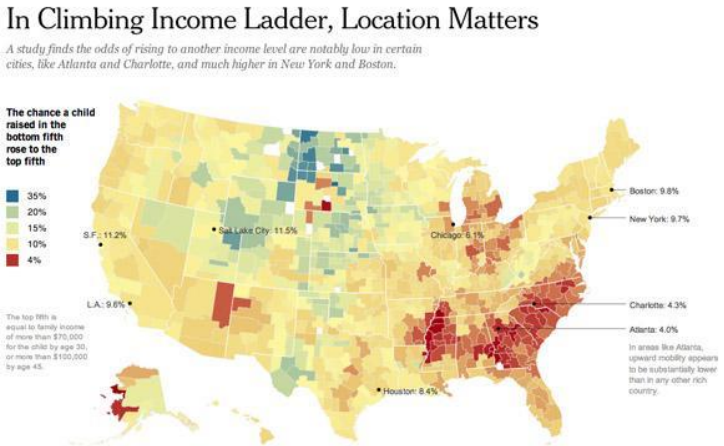
	HTML5	JAVA	FLASH	SOCIAL MEDIA	WEB/ COMMENT	SURVEY/ DATABASE	ADOBE/ GOOGLE APPS	OTHERS
Online Only	4%	4%	1%	0%	6%	7%	7%	17%
Hybrid	8%	2%	6%	4%	3%	3%	5%	10%

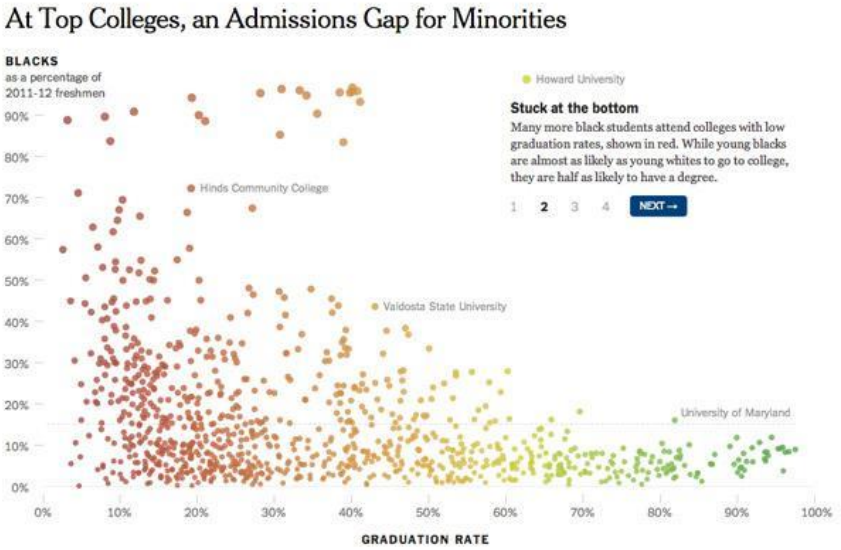
jdna journalism digital news archive

rji DONALD W. REYNOLDS JOURNALISM INSTITUTE
Missouri School of Journalism University of Missouri

Why are they worth preserving?

- What is the value of newsroom digital preservation?
  - Public Service
  - Posterity





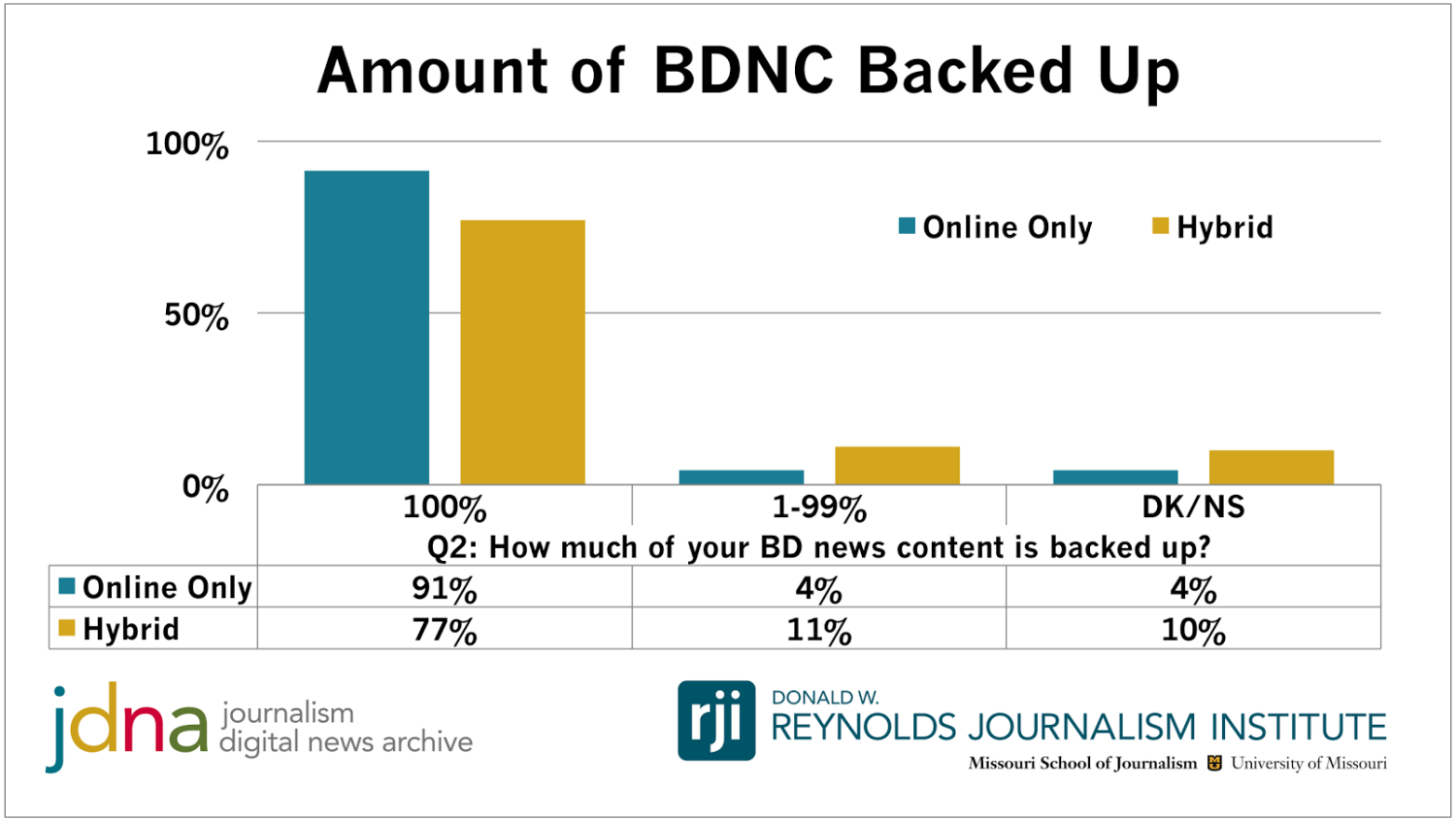
Imgs: http://nyti.ms/1h9NkB0

Intro to Newsroom Preservation

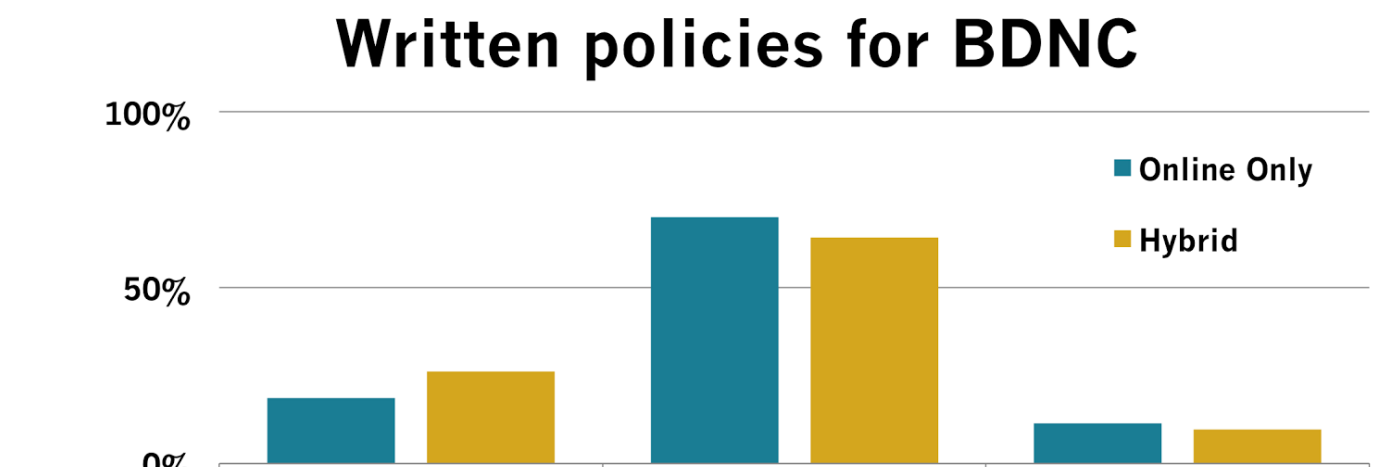
Concepts:

- What is being archived?
- What are the maintenance challenges?

Survey Results



Survey Results



Q8: Does your news org. have written policies for managing BD materials?

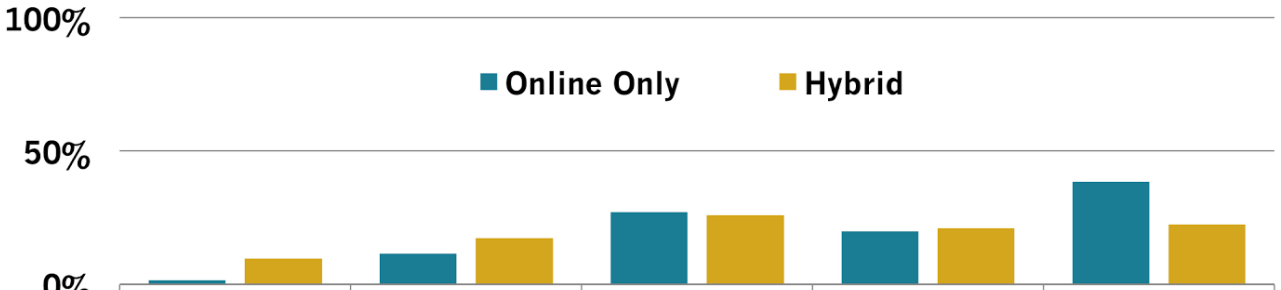
	Yes	No	DK/NS
Online Only	19%	70%	11%
Hybrid	26%	64%	10%

jdna journalism digital news archive

rji DONALD W. REYNOLDS JOURNALISM INSTITUTE
Missouri School of Journalism University of Missouri

Survey Results

Insufficient policy as threat to BNDC



Q13c - Rank lack of or insufficient policy or plan for preservation as a threat

	Great Threat	Threat	Moderate Threat	Less Threat	No Threat
Online Only	1%	11%	27%	20%	39%
Hybrid	10%	17%	26%	21%	22%

jdna journalism digital news archive

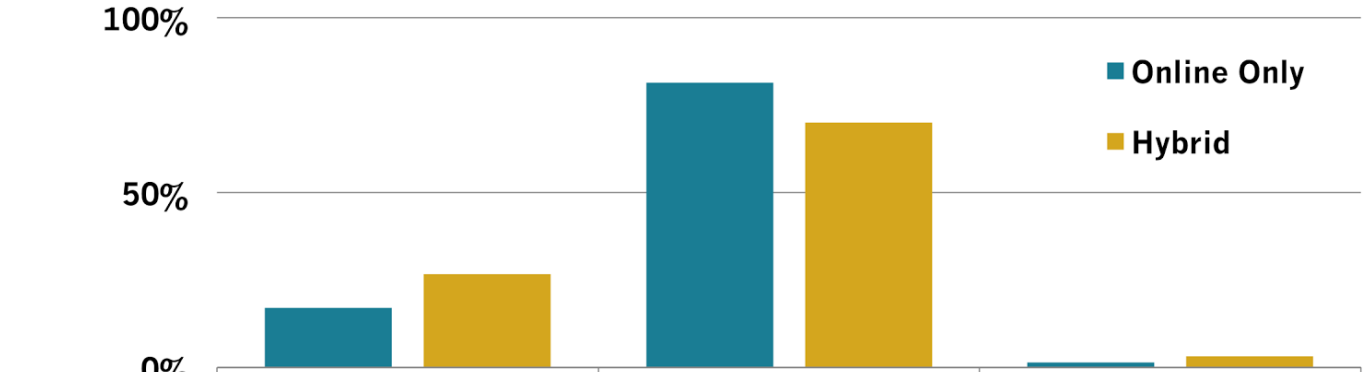
rji DONALD W. REYNOLDS JOURNALISM INSTITUTE
Missouri School of Journalism University of Missouri

Current Status

Preservation Practice

- What is the current approach to archiving in newsrooms?
  - Survey Results
  - Anecdotal

Survey Results



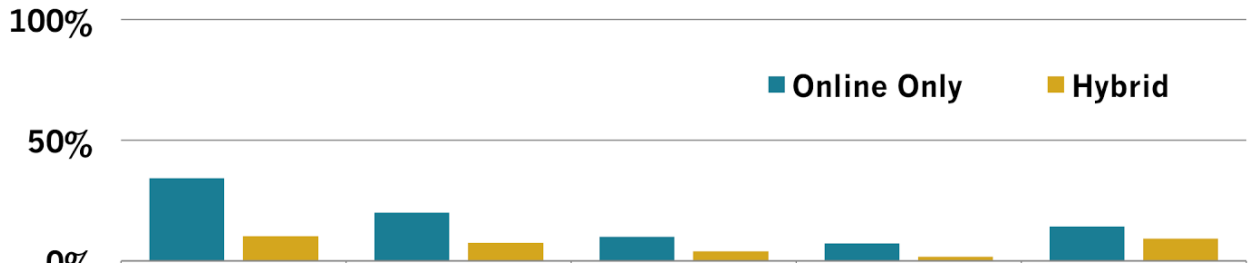
Q5c: Have you ever experienced a significant loss of news content?	Yes	No	DK/NS
Online Only	17%	81%	1%
Hybrid	27%	70%	3%

jdna journalism digital news archive

rji DONALD W. REYNOLDS JOURNALISM INSTITUTE
Missouri School of Journalism University of Missouri

Survey Results

BDNC not going to memory institutions

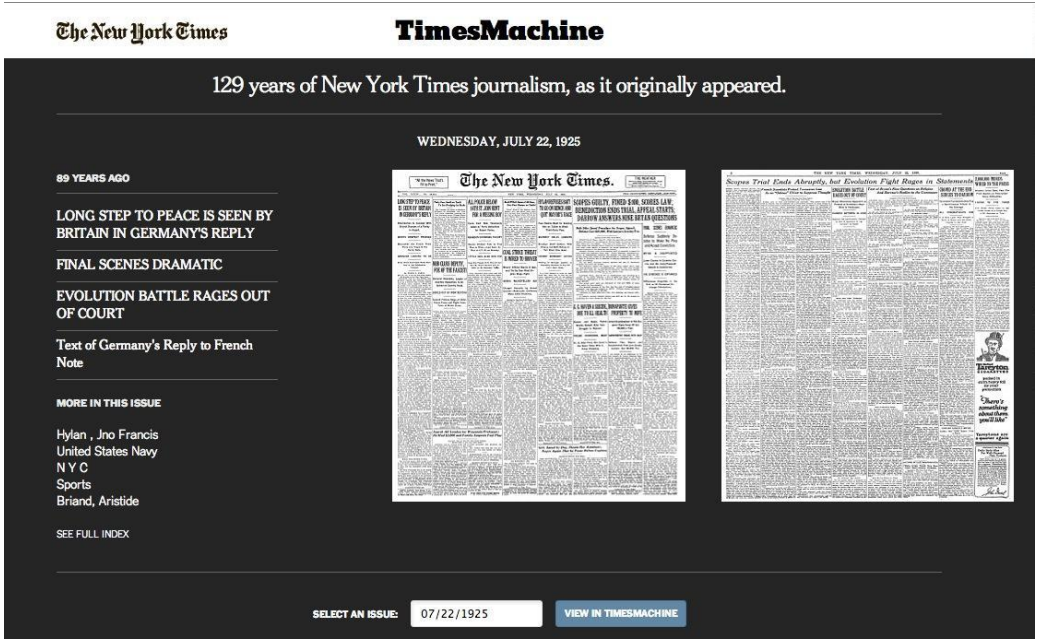


	Specify	Nothing Particular	Don't Know Inst.	Don't Want to Give	DK/NS
	Q6b: What is the primary reason you do not use a memory institution?				
Online Only	34%	20%	10%	7%	14%
Hybrid	10%	7%	4%	2%	9%

jdna journalism digital news archive

rji DONALD W. REYNOLDS JOURNALISM INSTITUTE
Missouri School of Journalism University of Missouri

What are large newsrooms doing?



Imgs: http://bit.ly/1nAeCnL
http://timesmachine.nytimes.com/

What are small newsrooms doing?

ProQuest | ProQuest Historical Newspapers™

Brooklyn Public Library

BROOKLYN DAILY EAGLE 1841-1902 Online™

Imgs: http://bit.ly/1wZBZXo

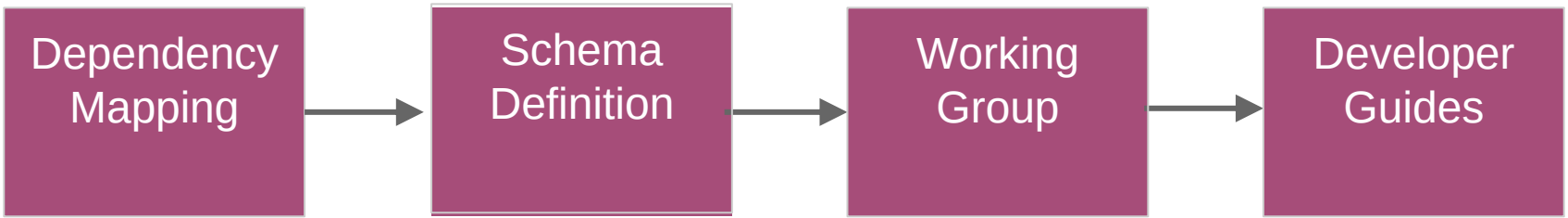
Proposed Practice

Designathon Plan

Develop a preservation program for newsrooms

Img: Anne Wooton

Some Outcomes



News Apps Model : Sheet1

Category	Artifacts	Attributes	Archival Requirement
Code	Front-end software, Back-end software, Data Code, Internal libraries	Actors, Documentation, Transformations, Versions, Decisions	Preserve
Data (Input)	Clean data, Raw data (copyright?), Metadata, Data Structure (field layout, data dictionary), UGC data, Live Data APIs, Reporting material,	Actors, Documentation, Transformations, Versions, Decisions	Preserve
Story (Output)	Narrative story, Links, Images/Audio/Multimedia, APIs we publish, UX, Viz Design, IA/Hierarchy/Taxonomy, Annotations, API Documentation	Actors, Documentation, Transformations, Versions, Decisions	Preserve
Infrastructure	The Internet, Browser, Server (OS, Language, Framework), Display APIs (external), Vendor libraries and dependencies, Platform (external hosting), Hardware, RDBMS/NoSQL, Bandwidth	Actors, Documentation, Decisions, Versions	Simulate, Emulate, Virtualize
Process	Code documented, Code "History" (git), Data Transformation, Data documentation, Cultural Context and Moment, Design Patterns, Edits of Story, Data sources (FOIA letters)	Actors, Documentation, Transformations, Versions, Decisions	Preserve
Response	Comments, Metrics, Awards, User Behavior, Logs, Inbound Links, Reaction/Media Coverage, Tweets, Impact, Intermediate Outcomes	Actors, Decisions	Preserve

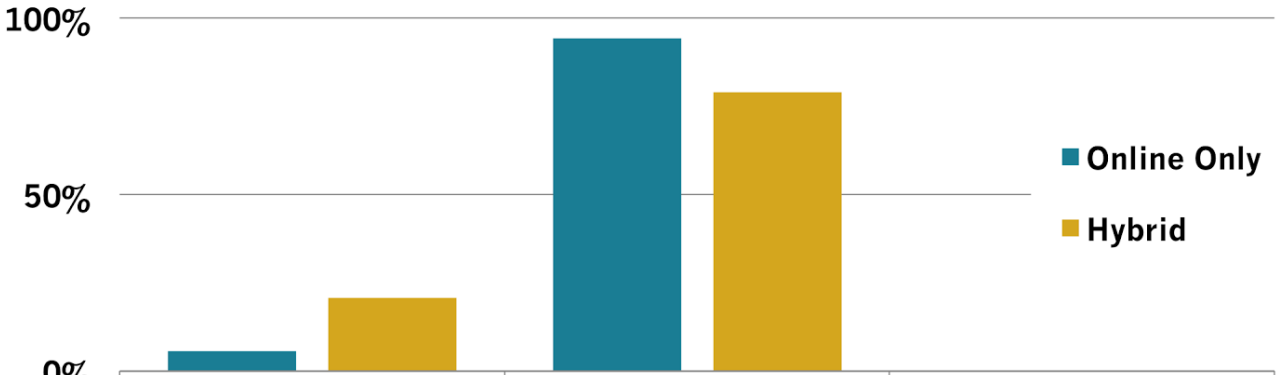
http://bit.ly/1nBMAbD | https://github.com/propublica/newsappmodel

Best/Better Practices

- Recommendations to news apps developers?
- Schema for cataloging and classifying news apps?
- Resource describing apps and their dependencies?

Survey Results

Does org. have a news librarian?



	Yes	No	DK/NS
	Q15c: Do you have a news librarian or equivalent position in your newsroom?		
Online Only	6%	94%	0%
Hybrid	21%	79%	0%

jdna journalism digital news archive

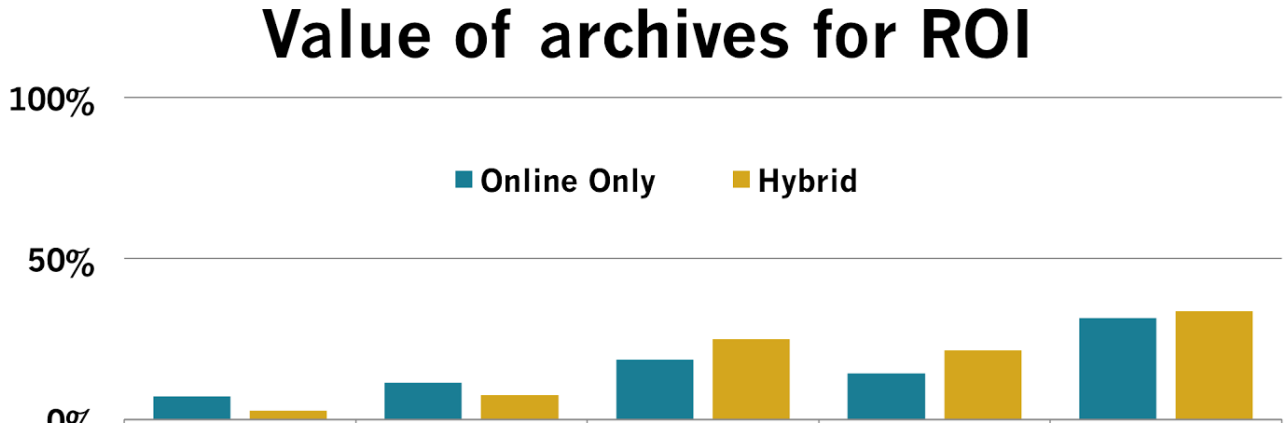
rji DONALD W. REYNOLDS JOURNALISM INSTITUTE
Missouri School of Journalism University of Missouri

Best/Better Practices

- Engaging partner communities in the process?
- Monetizing the archive?

Img: Banner Blindness Study, http://bit.ly/1nAwrmt

Survey Results



Q12d: How valuable is having access to your archives for producing good RIO?

	Not at all	Slightly Valuable	Moderately Valuable	Valuable	Very Valuable
Online Only	7%	11%	19%	14%	31%
Hybrid	3%	8%	25%	21%	34%

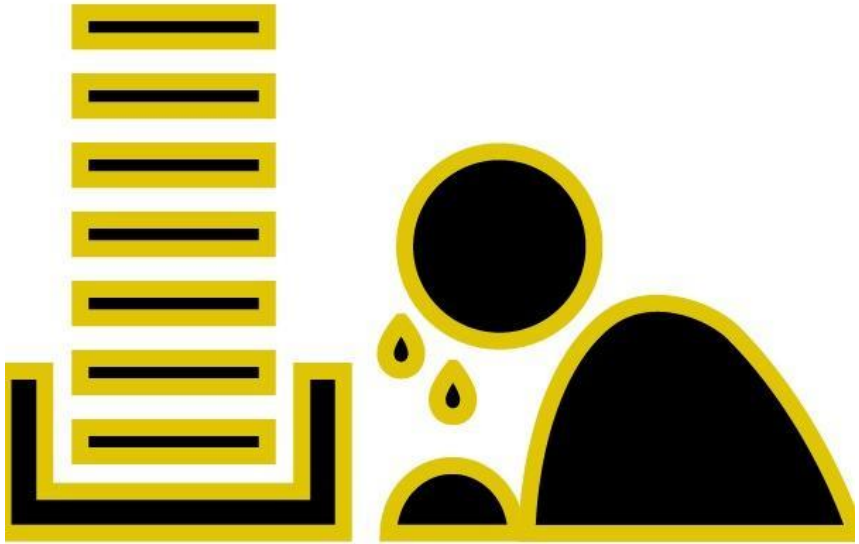
jdna journalism digital news archive

rji DONALD W. REYNOLDS JOURNALISM INSTITUTE
Missouri School of Journalism University of Missouri

Flexible Practice

Flexible Practices

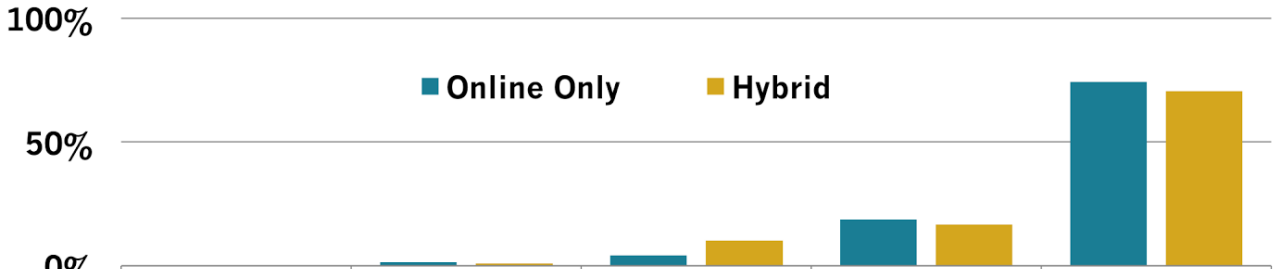
- What is an archive?
  - Values?
  - Threats?



Img: Result for “archive” keyword search in the Noun Project, correlates with “overworked”
http://bit.ly/1pc4T5j

Survey Results

Value of archives for quality journalism



Q12c: How valuable is having access to your archives for producing quality journalism?

	Not at all	Slightly Valuable	Moderately Valuable	Valuable	Very Valuable
Online Only	0%	1%	4%	19%	74%
Hybrid	0%	1%	10%	17%	71%

jdna journalism digital news archive

DONALD W. REYNOLDS JOURNALISM INSTITUTE
Missouri School of Journalism | University of Missouri

Institution-Level Scaling

- Can we accommodate all papers and preservation needs?

Img: http://bit.ly/1rAmBTB

Institutional Cases

large newsroom
http://nyti.ms/1h9NkB0

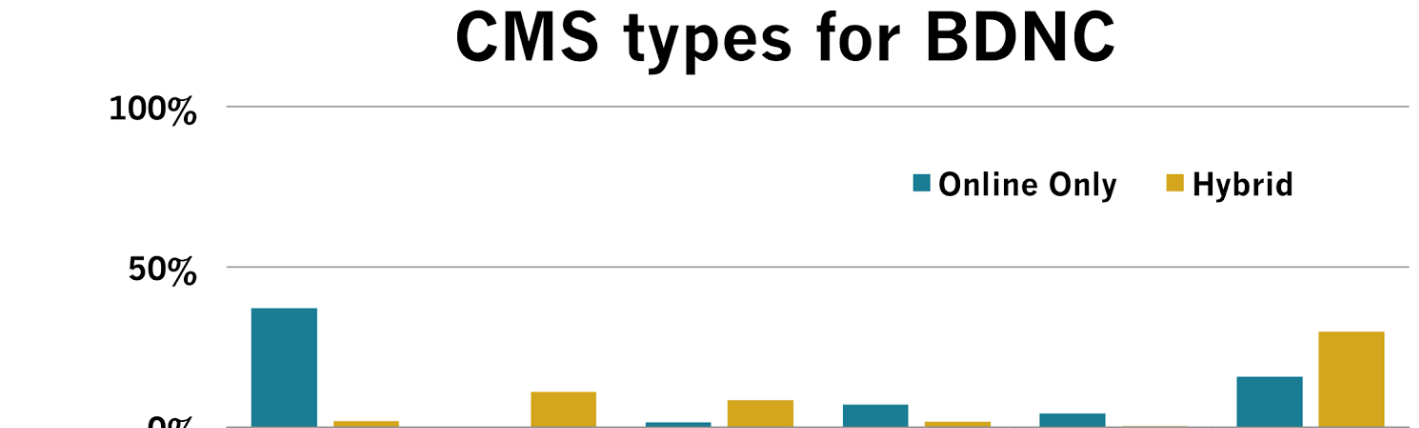
small newsroom
http://bklyn.newspapers.com/

Tools of Entrapment

- News orgs rely heavily their CMSes
- Need public/private partnerships for buy in around better preservation practice
- NYTimes Scoop

Img: http://bit.ly/1kbuxZ2 | http://nyti.ms/1rAUvHF

Survey Results



Q4bbOth: What specific CMS system do you use?

	WORDPRESS	TOWN NEWS	SAXOTECH	DRUPAL	JOOMLA	OTHER
Online Only	37%	0%	1%	7%	4%	16%
Hybrid	2%	11%	8%	2%	0%	30%

jdna journalism digital news archive

rji DONALD W. REYNOLDS JOURNALISM INSTITUTE
Missouri School of Journalism University of Missouri

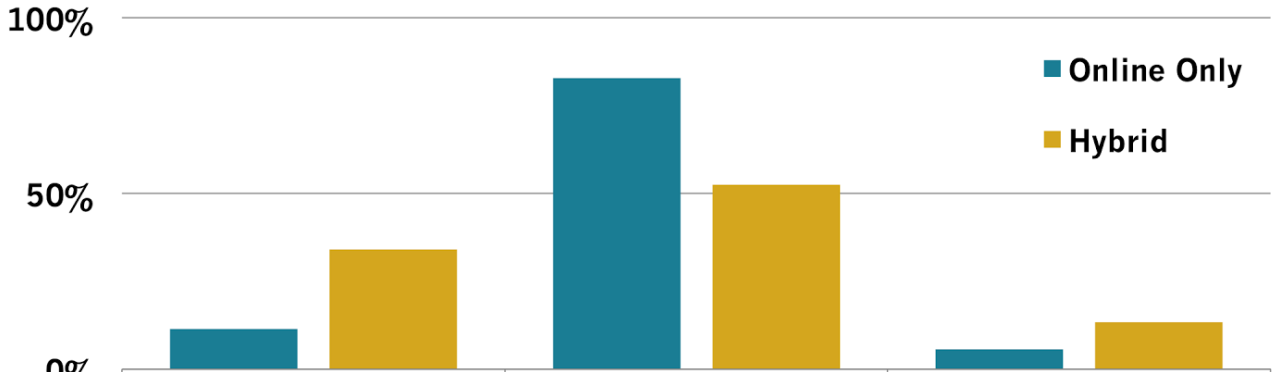
Value of the Archive

- How are newsroom archives quantifiably valuable?
- To what extent is monetization an option in the newsroom?

Img: http://bit.ly/1nejEoE

Survey Results

Revenue generation from vendors



	Yes–Specify	Nothing	DK/NS
	Q14a: Are you generating revenue through vendors such as ProQuest or Newsbank?		
Online Only	11%	83%	6%
Hybrid	34%	53%	13%

jdna journalism digital news archive

rji DONALD W. REYNOLDS JOURNALISM INSTITUTE
Missouri School of Journalism University of Missouri

Future Practice

Future Practice - Larger Papers

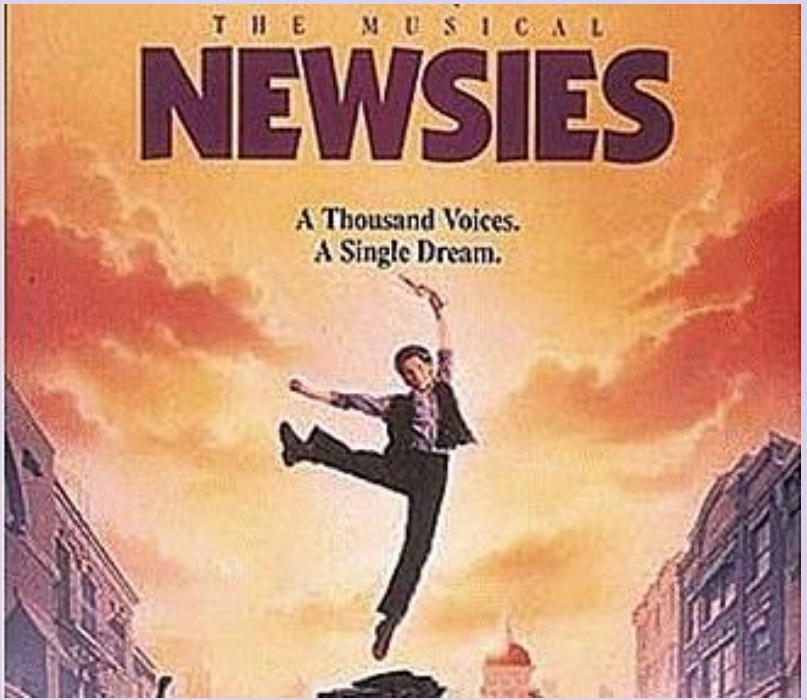
- Future media to preserve?
- Preserving leaks and breaches? Snowden files, Yanukovych leaked documents?

FREEDOM OF THE PRESS FOUNDATION

Img: https://pressfreedomfoundation.org/

Future Practice - Smaller Papers

- News co-ops for smaller papers?
- Enabling different media archives: audio, tv, broadcast
- Partnerships



Img: http://bit.ly/1jUUw6Q

Questions + Takeaways

- Designathon Takeaways:
  - Wiki: http://mzl.la/1pbwnYG
  - Blogpost: http://bit.ly/1nBMAbD
- Pop Up Archive:
  - Blogpost: http://bit.ly/1kbUisq
- Survey Results:
  - Blogpost: http://bit.ly/1wwlEt4
- Next Steps ...

Thank you!

Anne Wootton
@annewooton

Leslie Johnston
@lljohnston

Edward McCain
@e_mccain

Aurelia Moser
@auremoser