



# Missing links: The digital news preservation

**DISCONTINUITY**

Dorothy Carner  
University of Missouri-Columbia  
[carnerd@missouri.edu](mailto:carnerd@missouri.edu)

Edward McCain  
Donald W. Reynolds Journalism Institute  
University of Missouri-Columbia  
[mccaine@rjionline.org](mailto:mccaine@rjionline.org)

Frederick Zarndt  
IFLA Newspapers Section  
[frederick@frederickzarndt.com](mailto:frederick@frederickzarndt.com)



# Our smartphone-obsessed society will leave behind few digital Mona Lisas



... Numerous photo sharing sites have shut down in the last five years, many with very little warning. The popular travel photography site *Fotopedia* is being shut down this weekend.

... I'm under no illusion that our Instagram selfies, our Twitter and Facebook feeds, our YouTube videos, our WordPress blogs and our photos living on Flickr and similar hosting sites will be around in 100 years, let alone 50 or even 25 unless we make a strong effort to identify digital works of cultural significance.

... while not all works being created digitally necessarily merit preservation, ultimately, for those that do, we need a plan to vault them long term, and document methods for retrieving that data far in the future.

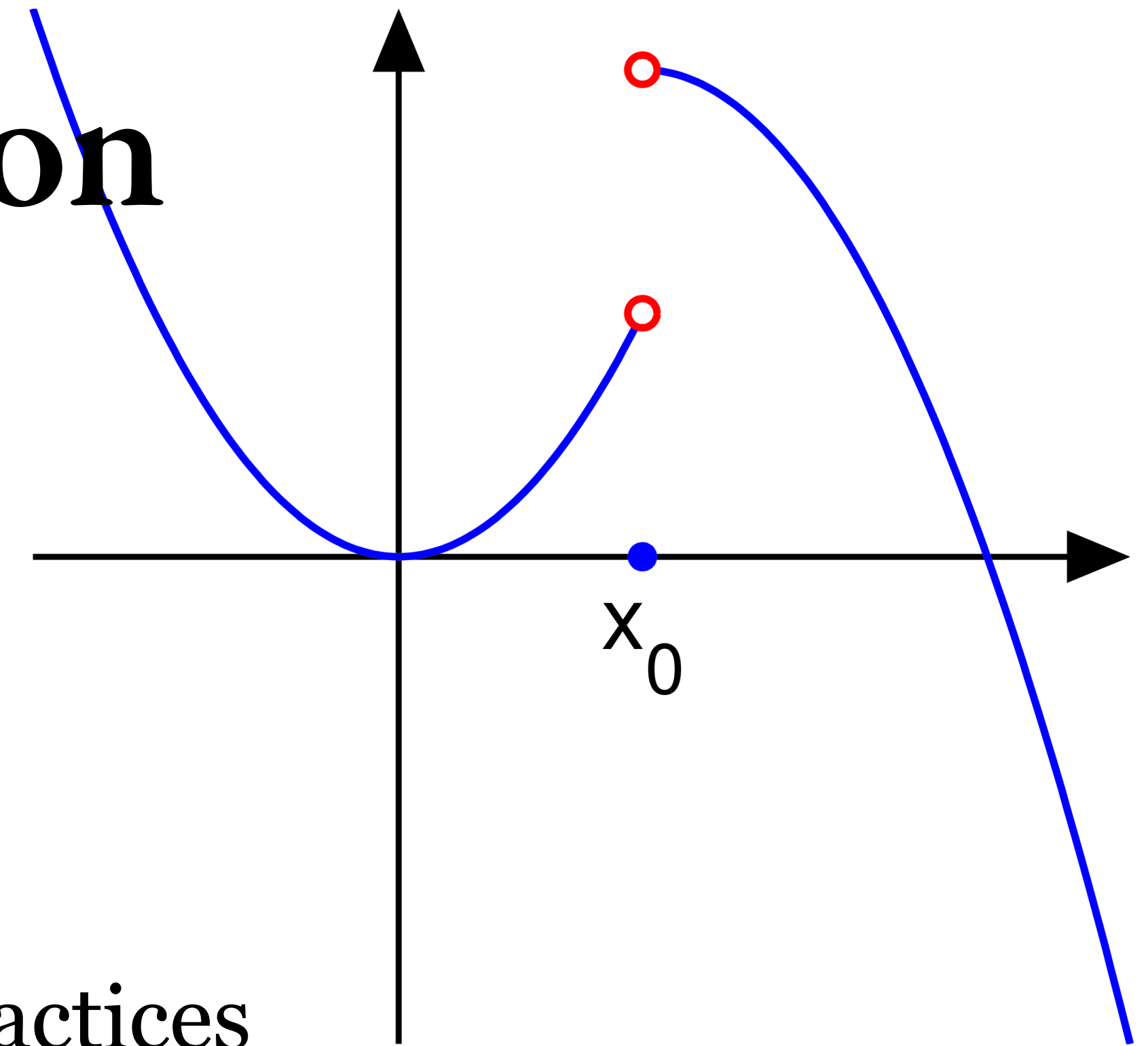
# Missing links?



- Legal deposit policy is (mostly) well developed for news**papers**
- Legal deposit procedures at cultural heritage organizations work well for news**papers**
- Legal deposit policy and procedures are in their infancy or non-existent for born digital news
- Question: How much born digital news — most of which will never see the printed page — is at risk of loss?

# The digital news preservation

## DISCONTINUITY



In this study we survey ...

- ... historical news preservation policies and practices
- ... born digital news preservation policies and practices of publishers
- ... cultural heritage organizations around the world about born digital news preservation policies and practices in their country



sideBAR

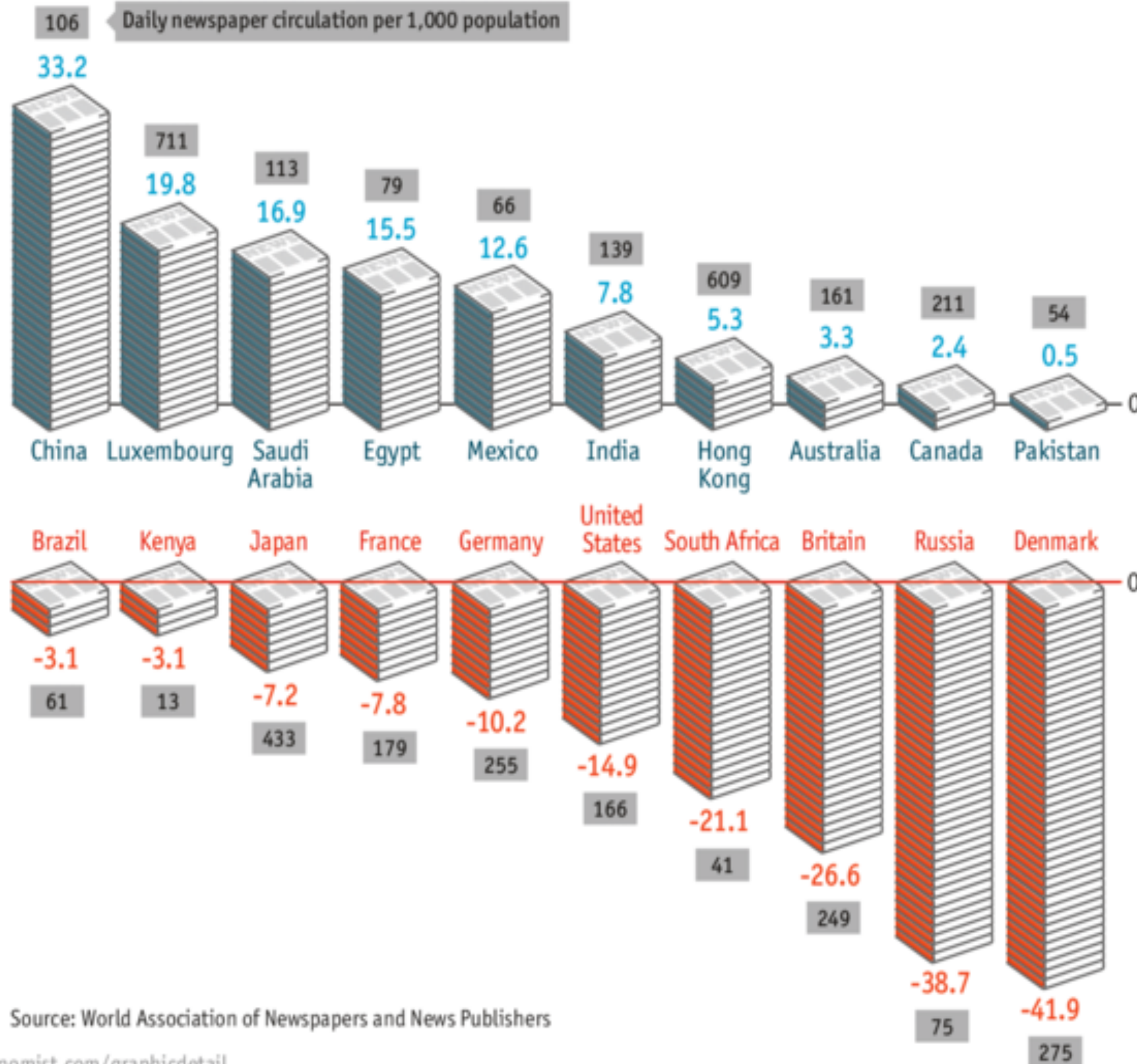
trends in

reading

**NEWS**

## Daily newspaper circulation

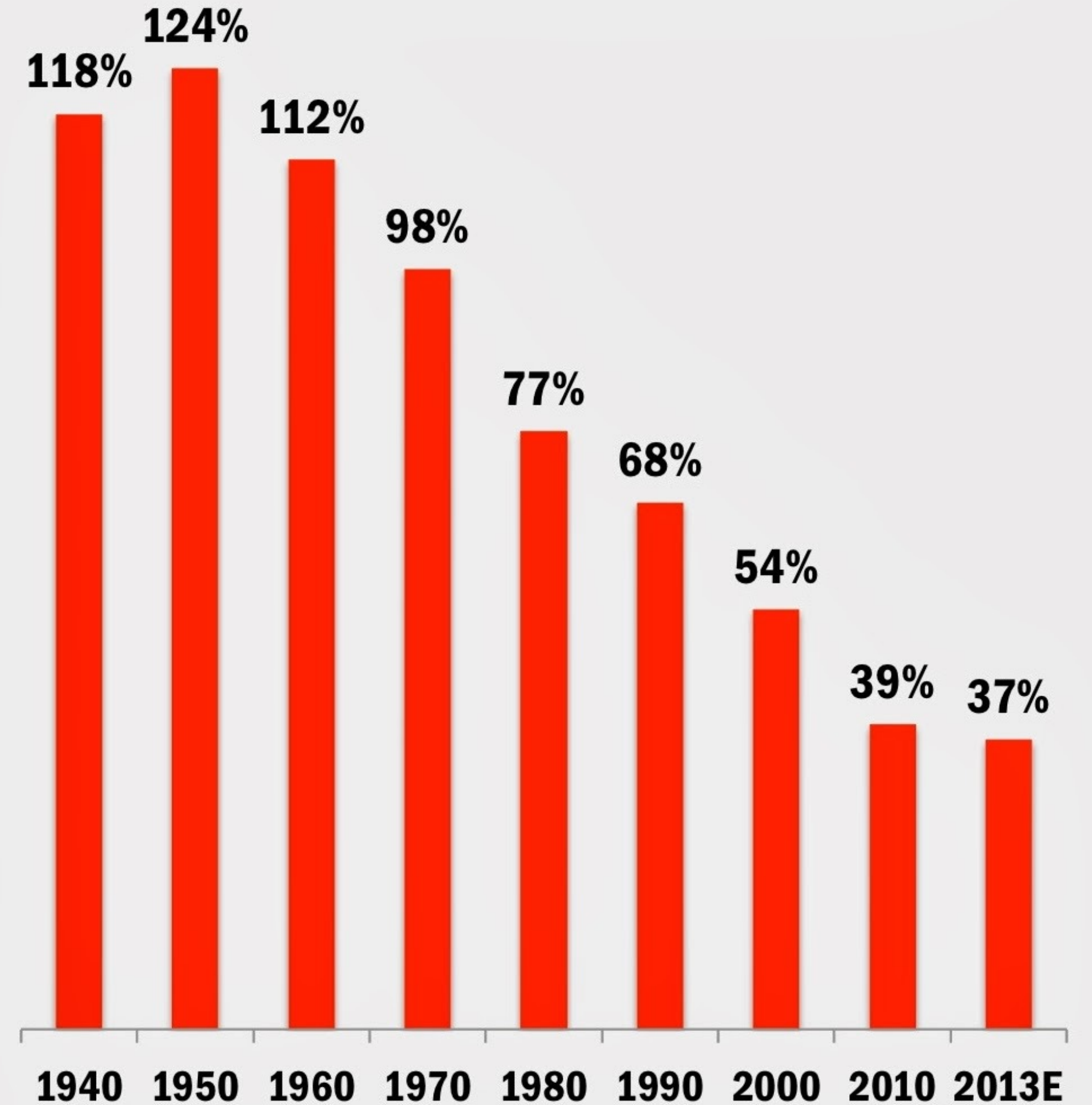
2008-12, % change



Source: World Association of Newspapers and News Publishers

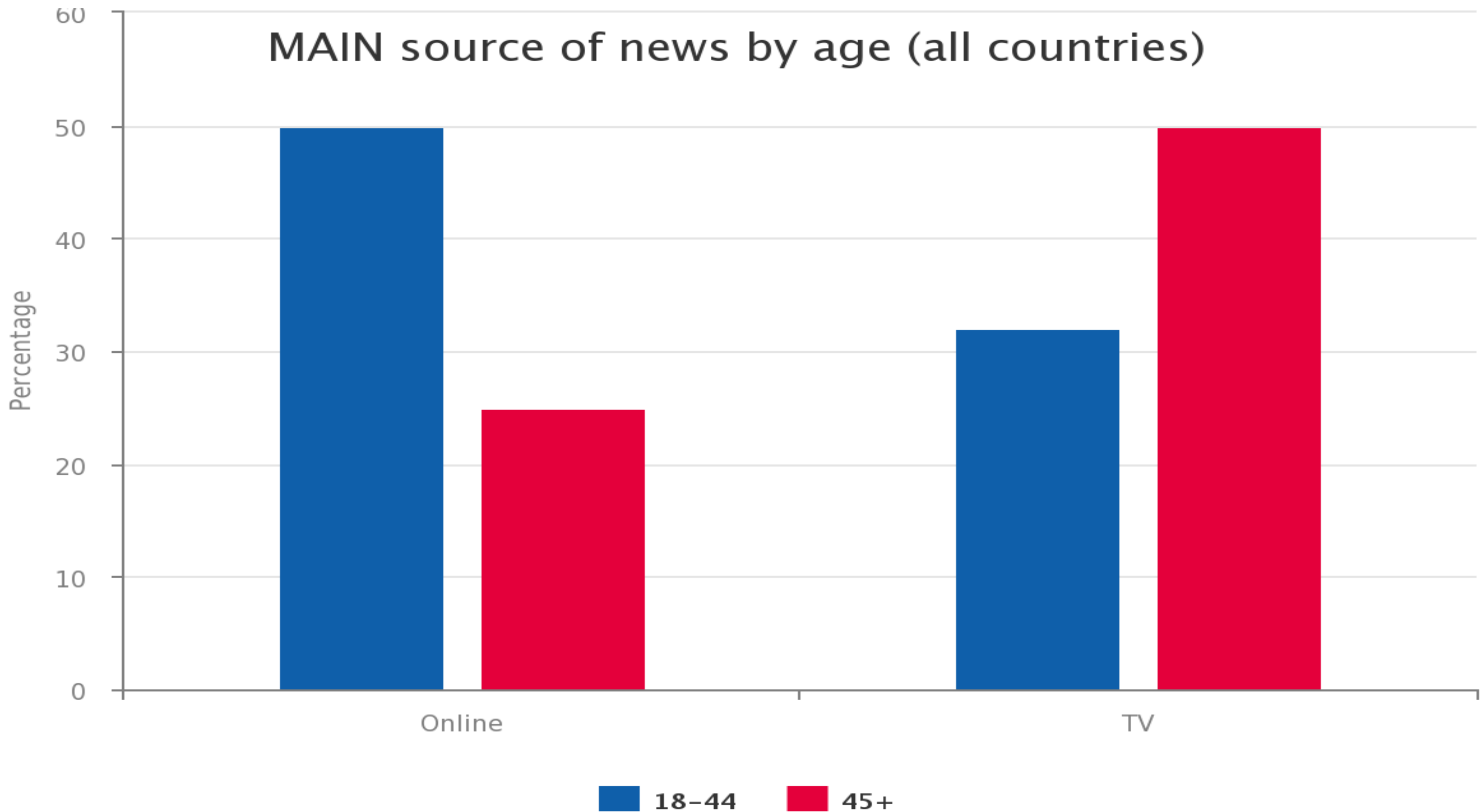
Economist.com/graphicdetail

## Percent of U.S. households taking weekday newspaper

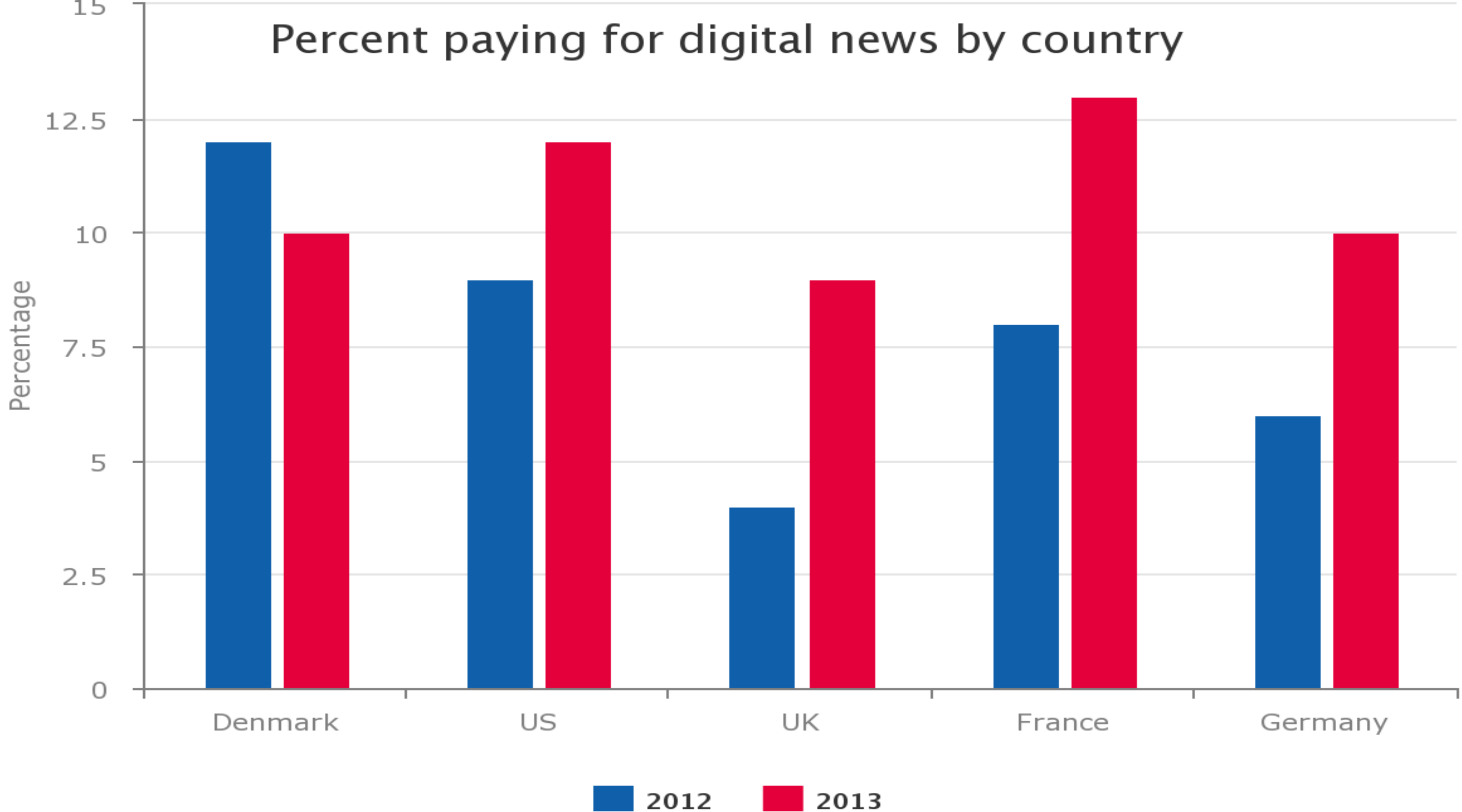


Sources: Editor & Publisher Yearbook and U.S. Census Bureau

# MAIN source of news by age (all countries)



# Percent paying for digital news by country





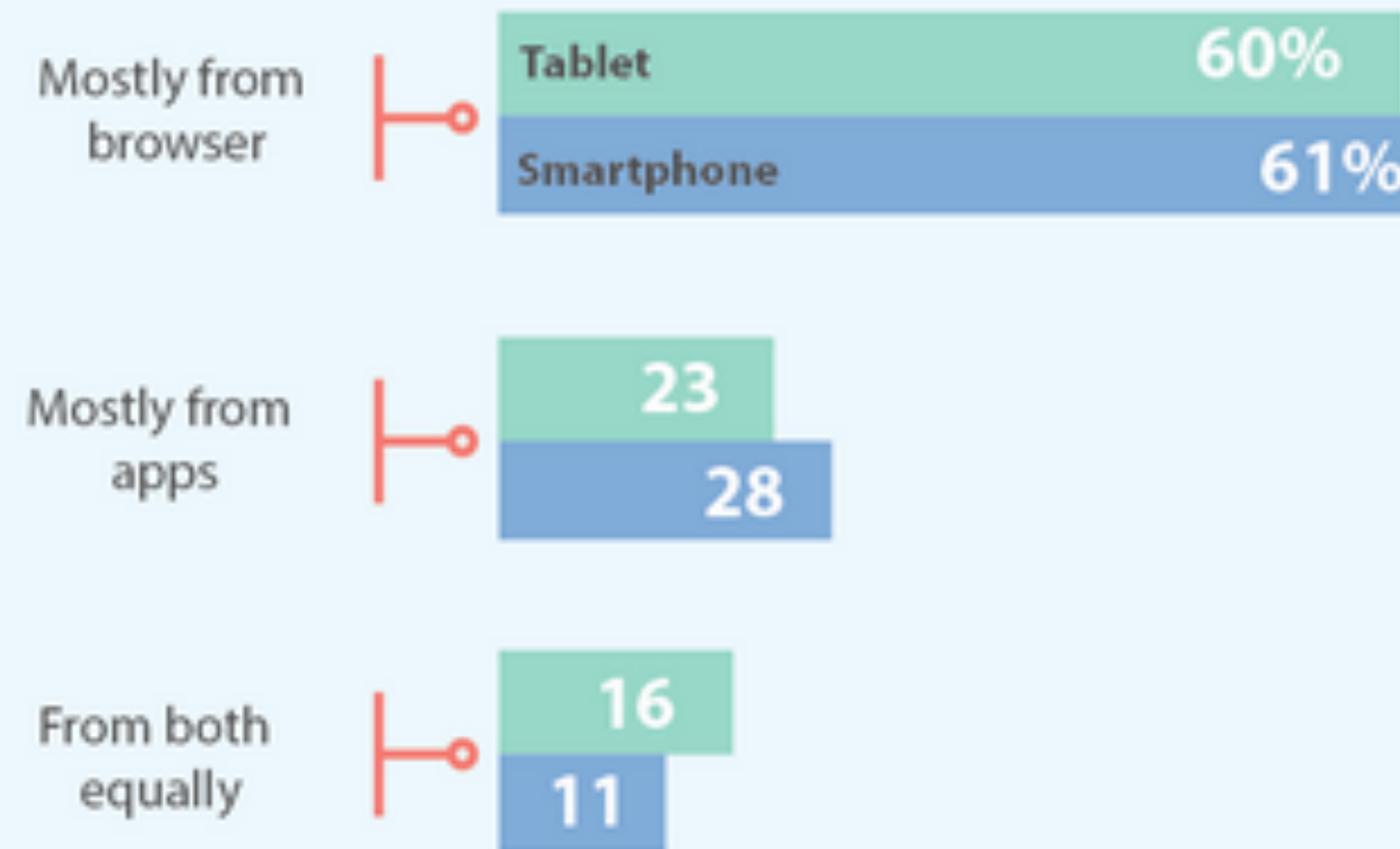
# The Washington Post

The Washington Post's Topicy, a mobile and visual news interface launched in September, 2013 was developed to rethink the news experience for mobile users' preferences.



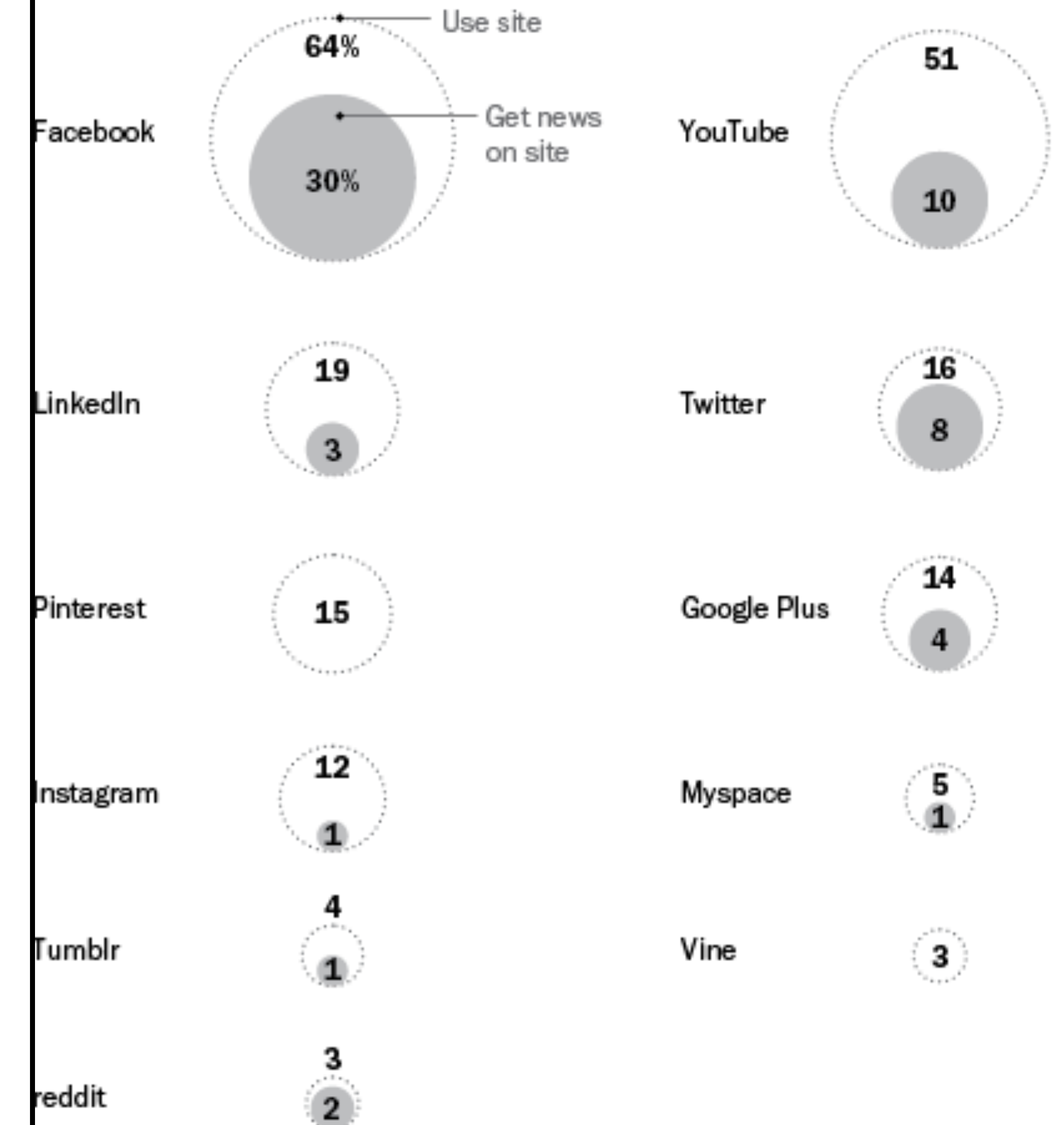
## More People Use a Mobile Browser for News

Percent of news consumers who get their news...



## Social Media as a Pathway to News: Facebook Leads the Way

Percent of U.S. adults who use each social networking site & percent of U.S. adults who get news from each social networking site



Note: The percent of U.S. adults who get news on Pinterest and Vine each amount to less than one percent.

Aug. 21-Sept. 2, 2013

PEW RESEARCH CENTER

Recruitment ADvisor. 3 Digital Trends for Publishers in 2014. Apr 2014. Accessed August 2014  
<http://blog.realmatch.com/trade-publishers/3-digital-trends-for-publishers-in-2014-infographic-part-1/>

THE FUTURE IS OF NEWS IS DIGITAL

Pew Research Journalism Project. News Use Across Social Media Platforms. Nov 2013.  
 Accessed August 2014 <http://www.journalism.org/2013/11/14/news-use-across-social-media-platforms/>



sideBAR

longevity  
of a

**BUSINESS**

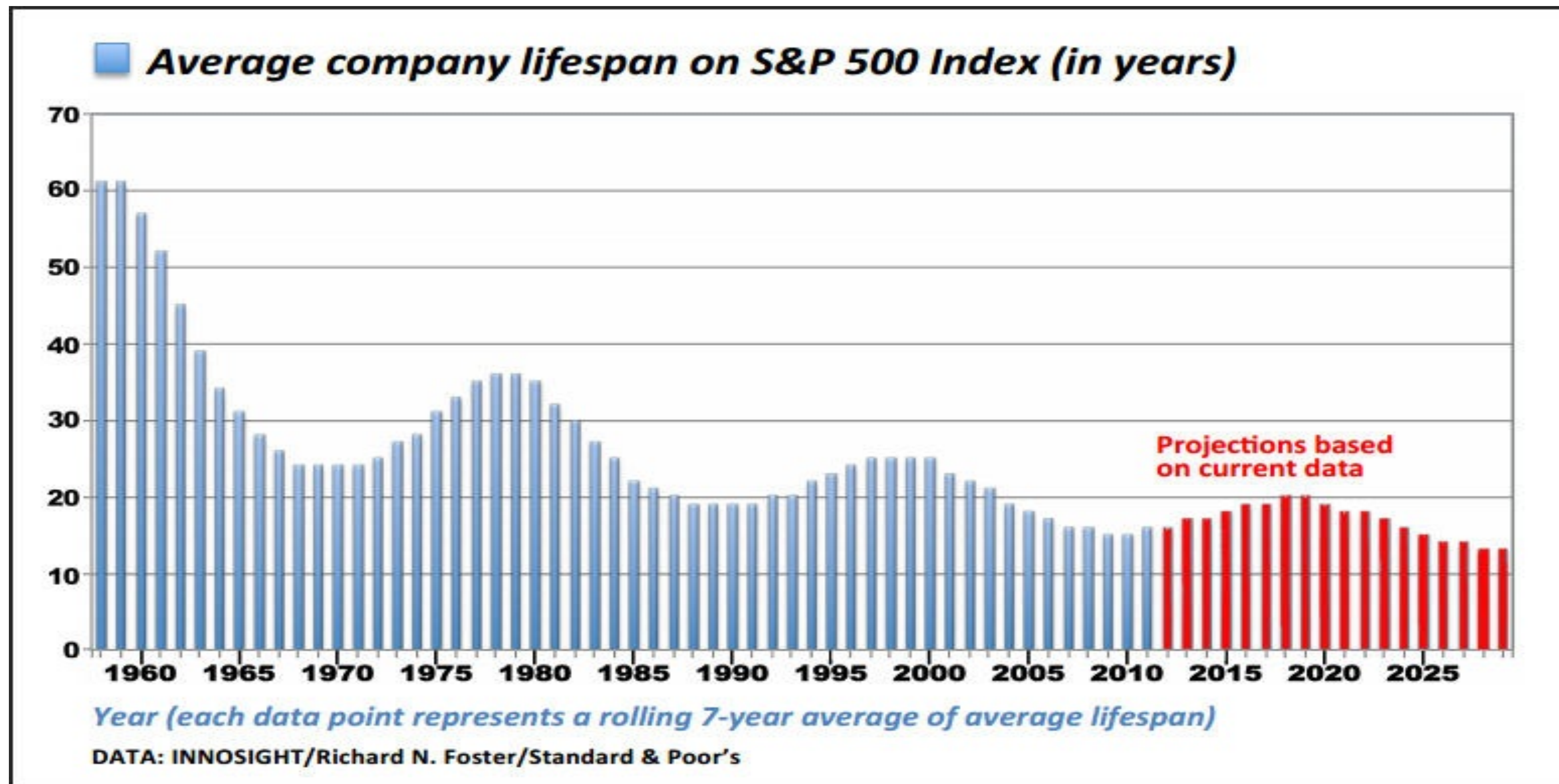
# Interesting fact



“...the big, solid companies, the pillars of the society we live in, seem to hold out for not much longer than an average of 40 years. And that 40-year figure, short though it seems, represents the life expectancy of companies of a considerable size.”

*Arie de Geus. The living company: Habits for survival in a turbulent business environment. Harvard Business Review Press. June 2002. p. 2.*

Graphic from *Creative destruction whips through corporate America*. Innosight Executive Briefing Winter 2012.



“..the life-expectancy of businesses in Standard & Poor’s 500 has fallen from 75 years in 1937 to just 15 years today.”

Paraphrased from Richard Foster and Sarah Kaplan. *Creative destruction: Why companies that are built to last underperform the market and how to successfully transform them*. McKinsey & Company. 2001.

## Lloyd's List

Lloyd's List 1734 : 280 years old [digital only since Dec 2013]



Seattle Post-Intelligencer

Seattle Post Intelligencer 1863 : 151 years old [digital only since Mar 2009]

## **The New York Times**

New York Times 1851 : 163 years old



Christian Science Monitor 1908 : 104 years old [ceased printing a daily paper Oct 2008, still prints a weekend edition]

online only

hybrid

Slate



Slate 1996 : 18 years old

Huffington Post 2005 : 9 years old

ProPublica 2007 : 7 years old

The Verge 2011 : 3 years old

Vox 2011 : 3 years old

PandoDaily 2012 : 1+ years old

online only

A national library is “specifically established by the government of a country to serve as the preeminent repository of information for that country.”

[[http://en.wikipedia.org/wiki/National\\_library](http://en.wikipedia.org/wiki/National_library)]



National Library of Finland 1640 : 374 years old



British Library 1973 / 1753 : 41 / 261 years old



Library of Congress 1800 : 214 years old



Russian State Library 1862 : 152 years old



sideBAR

digital  
newspapers  
USERS



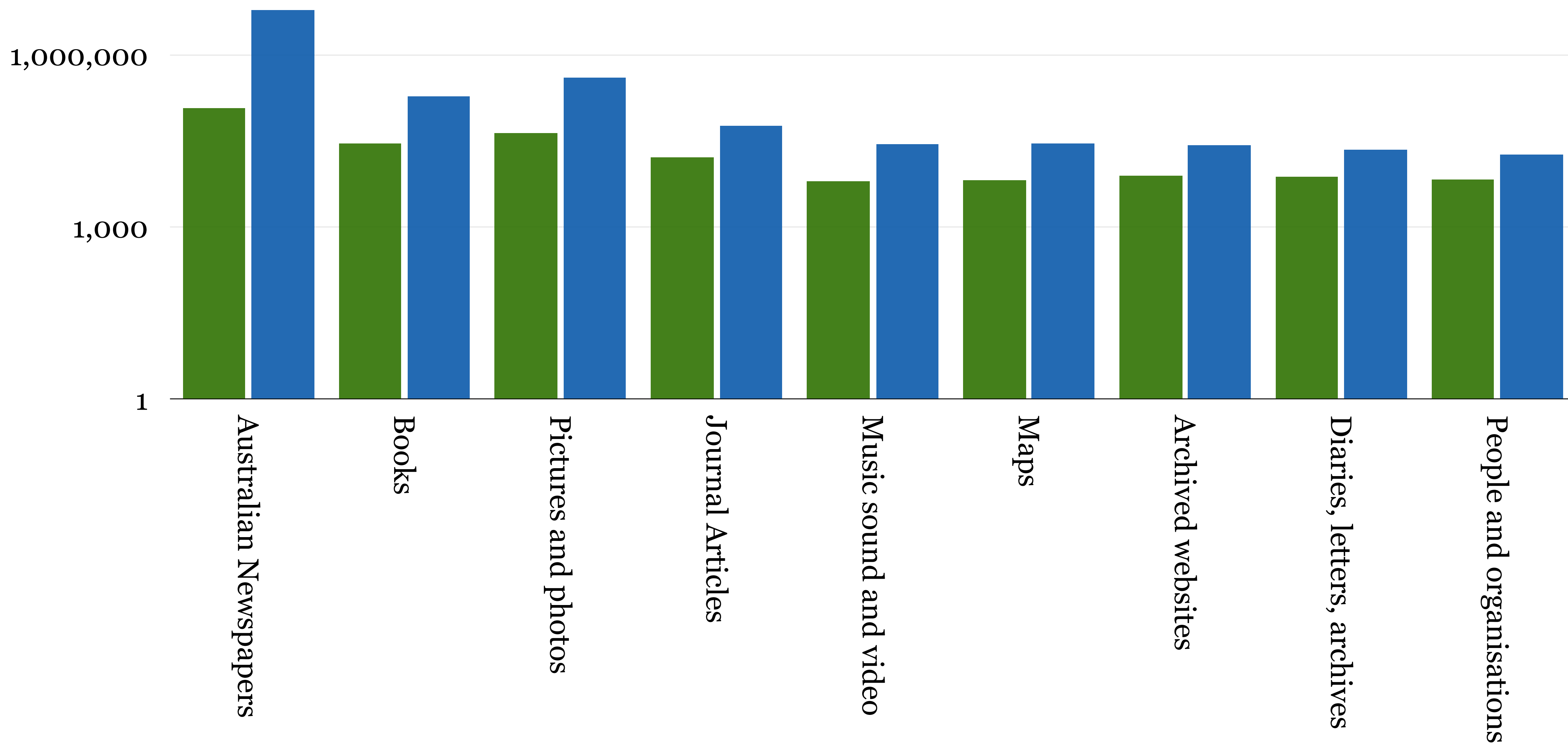


# Trove

# 2013 monthly averages

■ unique visits

■ page views



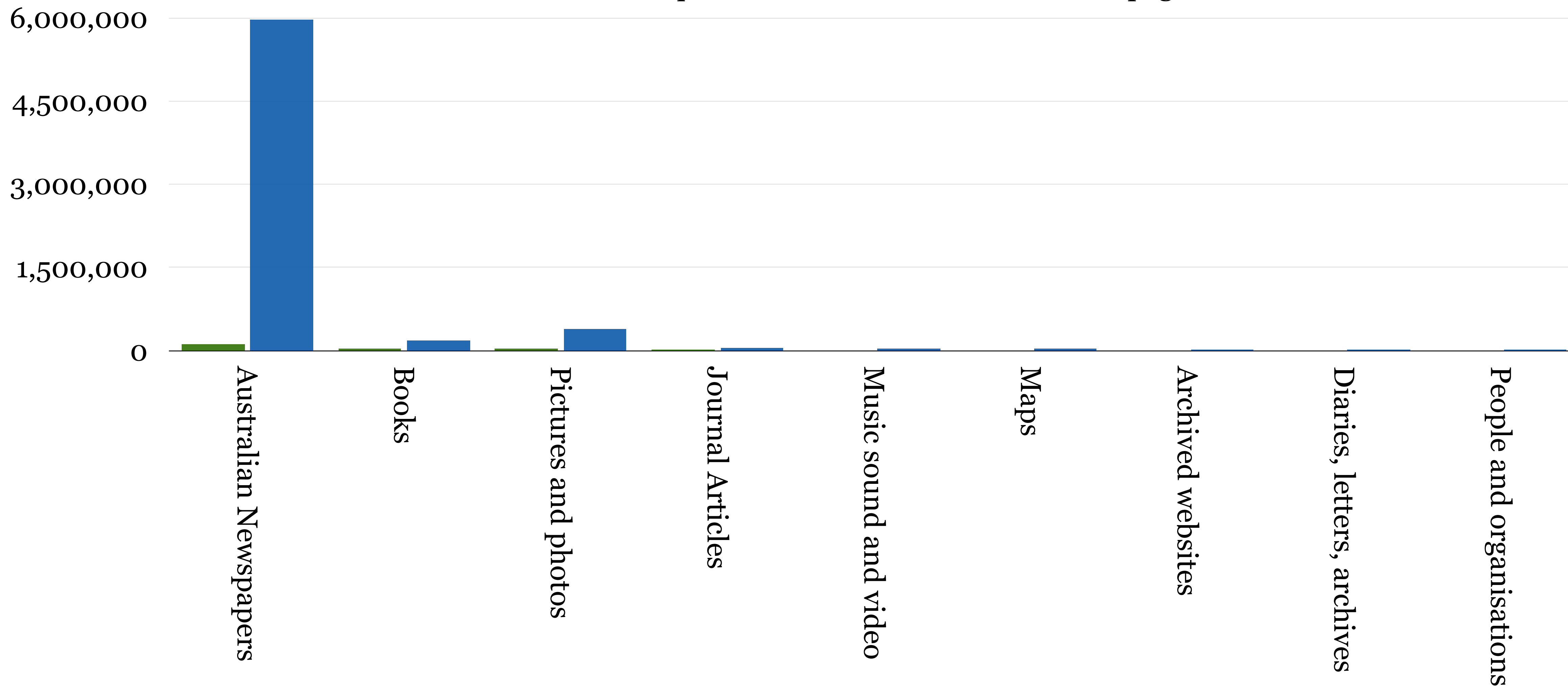


# Trove

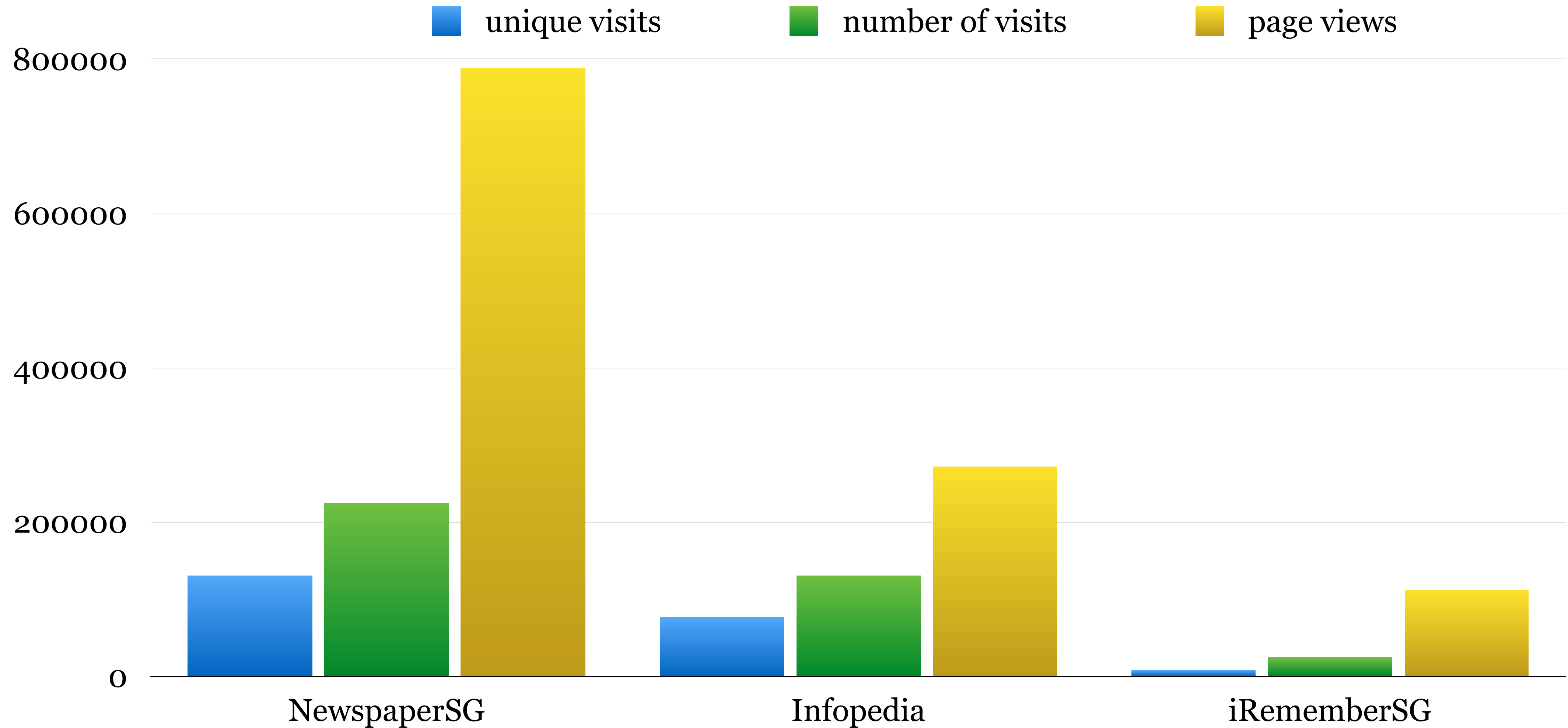
# 2013 monthly averages

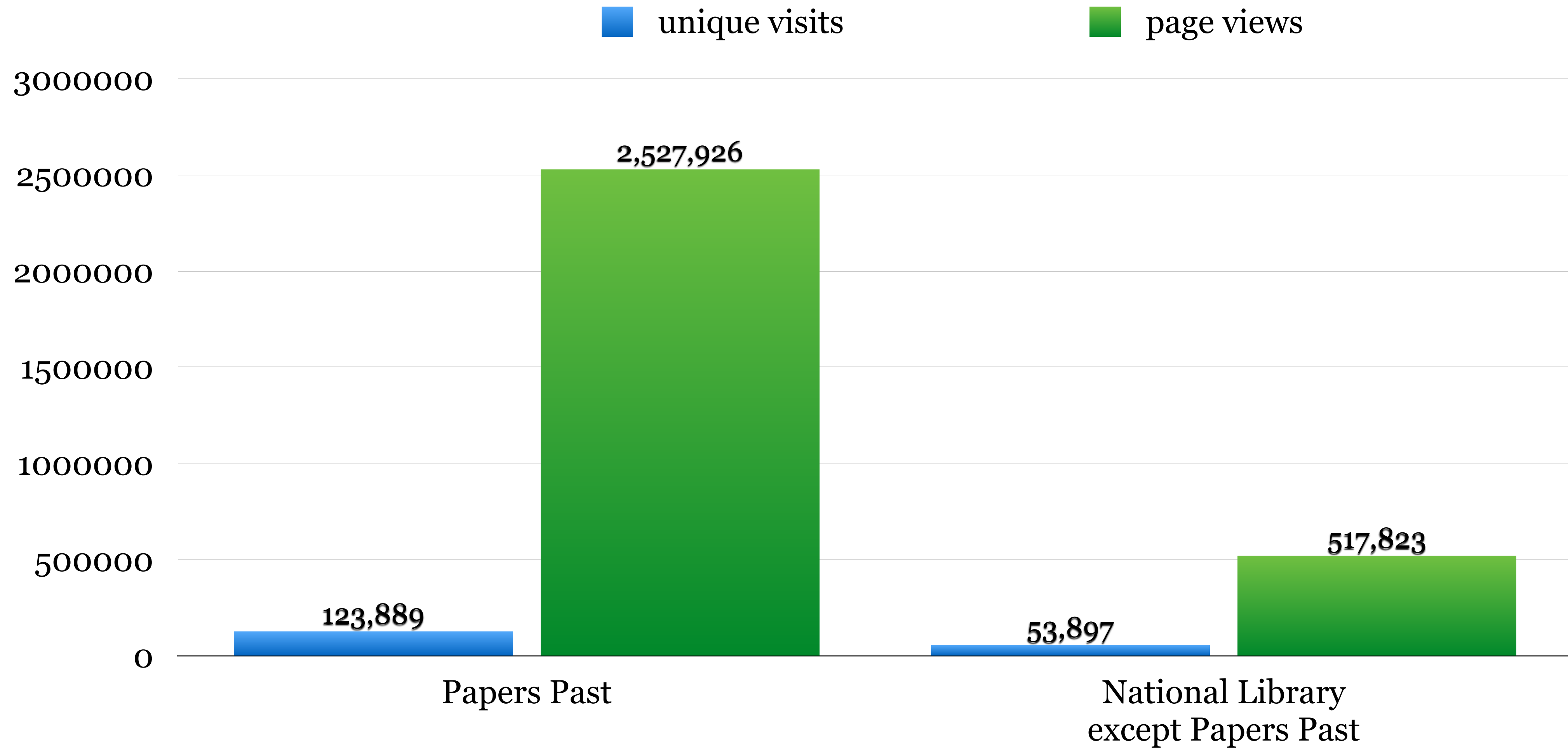
■ unique visits

■ page views



# 2013 monthly averages

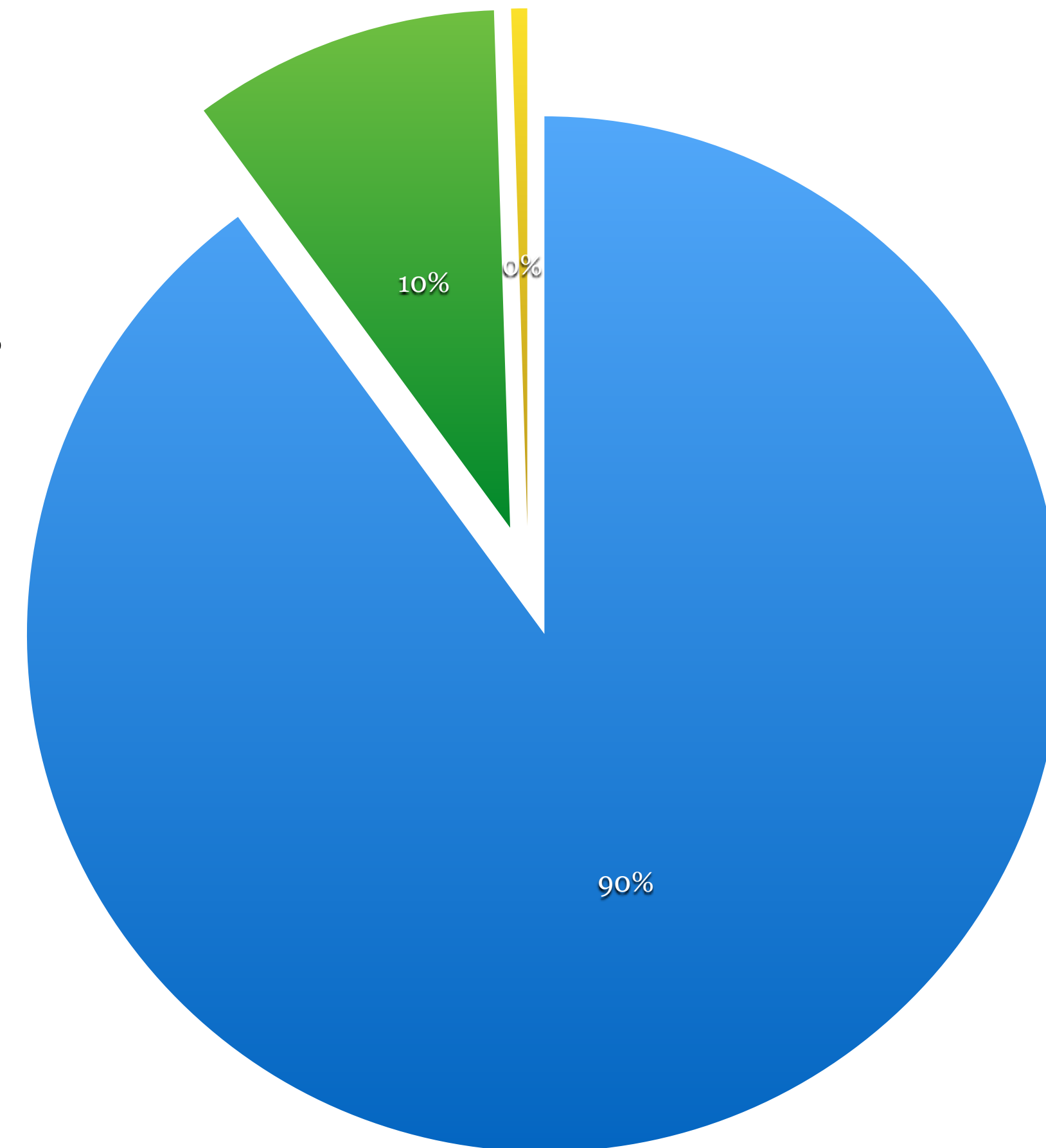






# 2013 monthly averages

- Historic Cambridge Newspapers (1846-1923)
- Cambridge City Directories (1848 - 1910)
- Cambridge Chronicle (August 2005 to present)



# Who uses library digital newspaper collections?



- National Library of Australia's 2012 Trove status report showed that ~50% of Trove users are family historians

PAPERSPAST

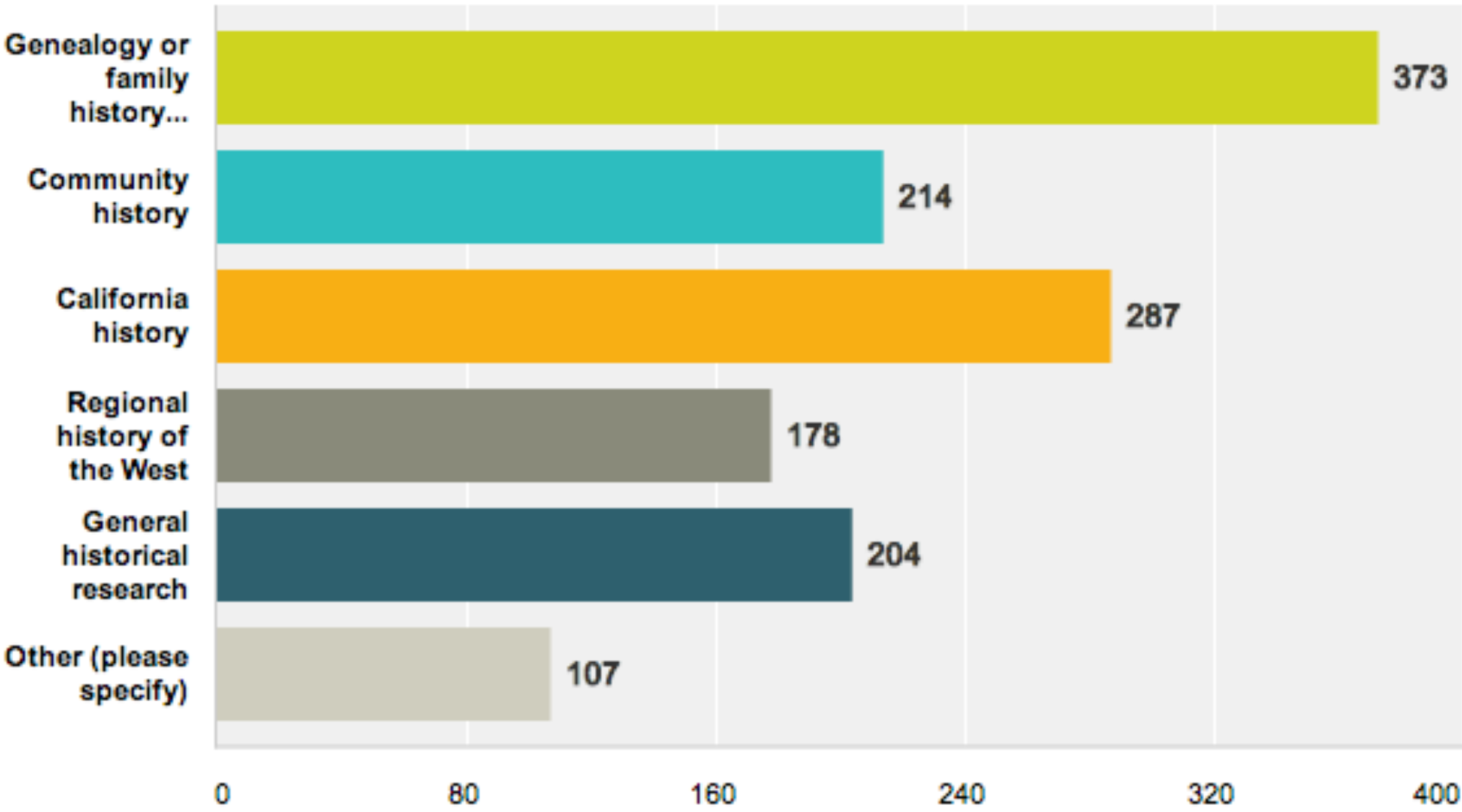
- National Library of New Zealand survey found that ~50% of PapersPast users are genealogists

- Utah Digital Newspapers learned that 72% of its users visit the collection for genealogical purposes in 2012 user survey

# Who uses library digital newspaper collections?

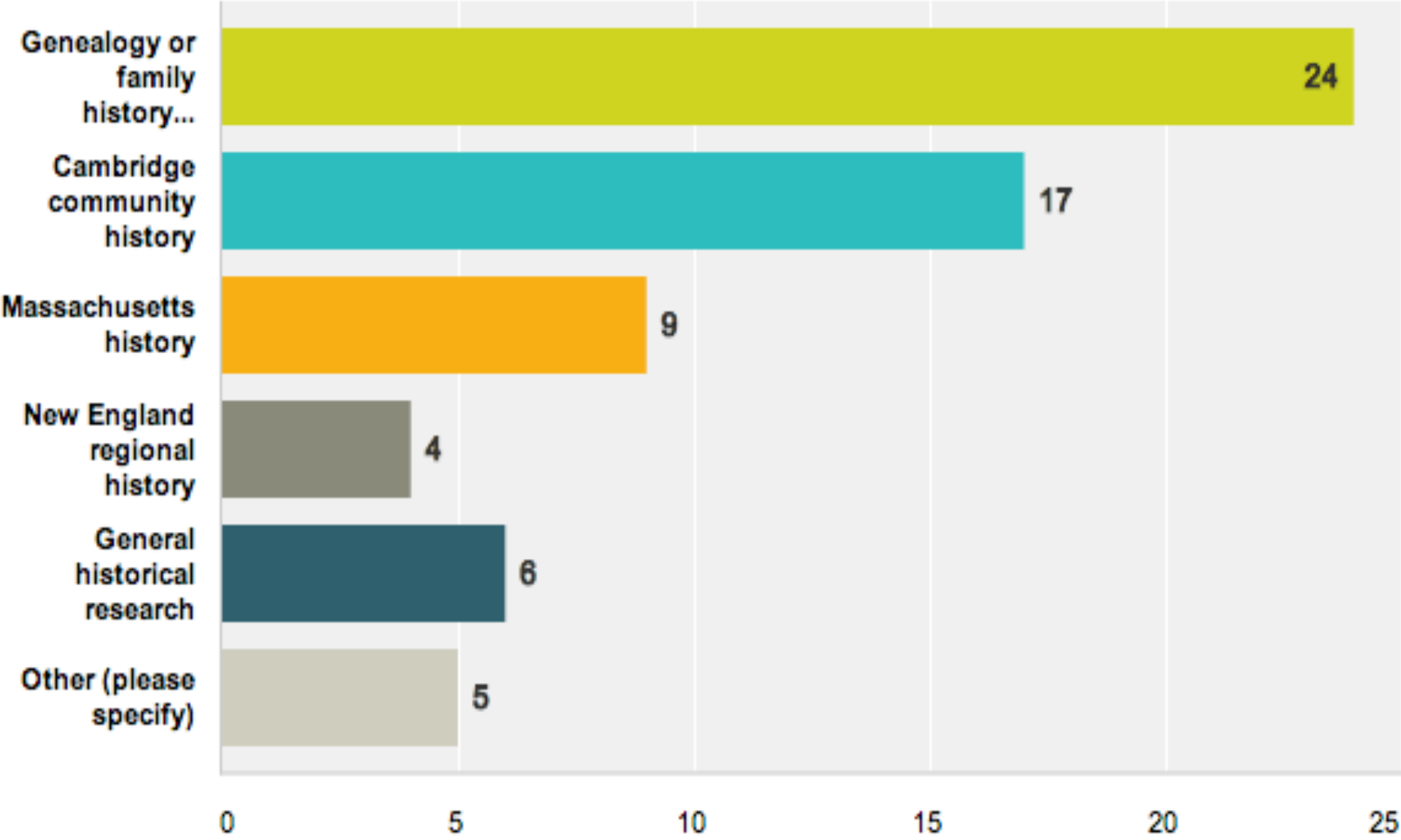
What are the main reasons you use CDNC (check all that apply)?

Answered: 604 Skipped: 0



What are the main reasons you use the Historic Cambridge Newspaper Collection (check all that apply)?

Answered: 30 Skipped: 2



END of

sideBAR





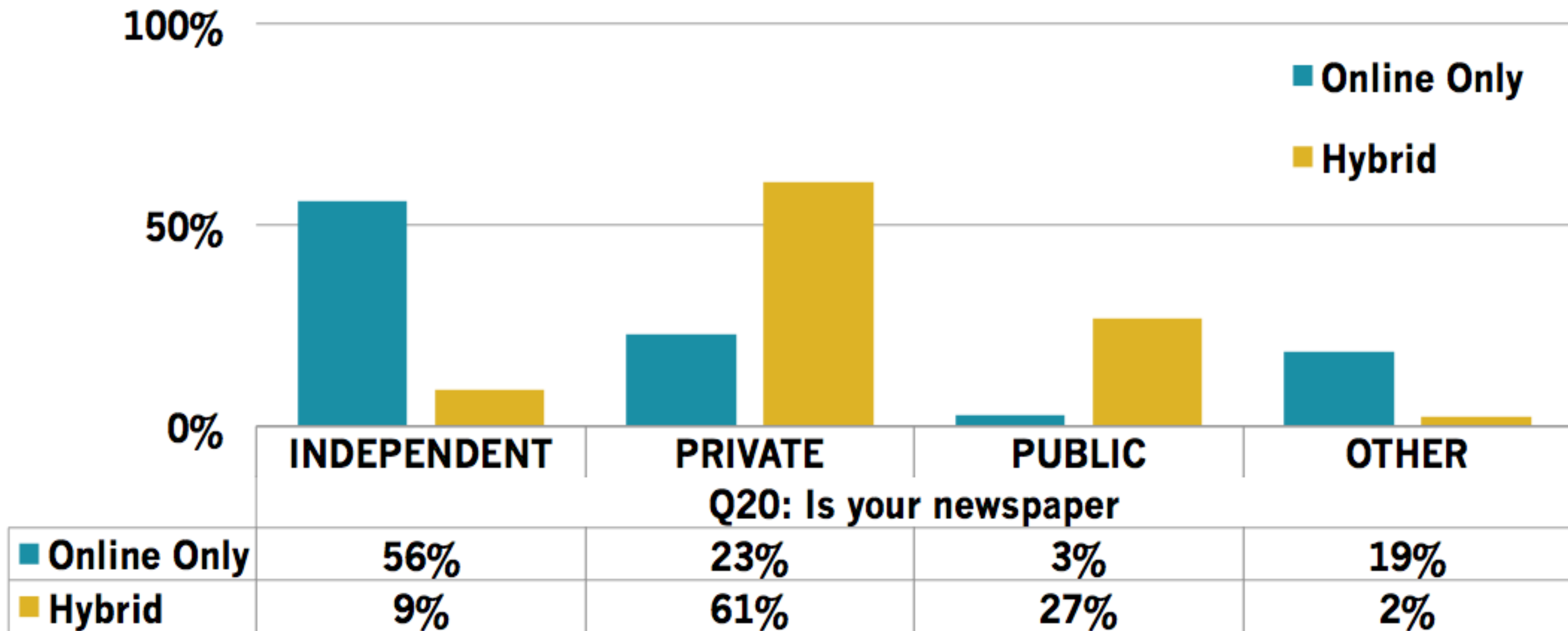


# Survey results

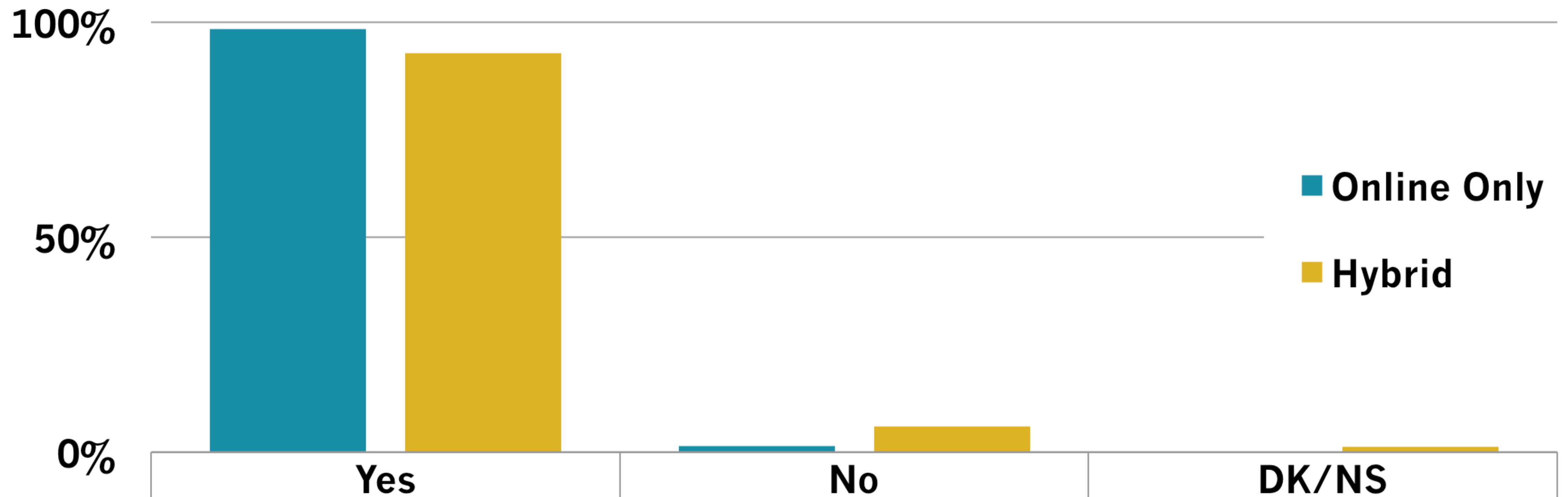
Reynolds Journalism Institute survey of 476 news organizations

- 406 “hybrid” organizations, producing both digital and print news
- 70 were “online only” organizations, producing only digital news

# Ownership type of news organization



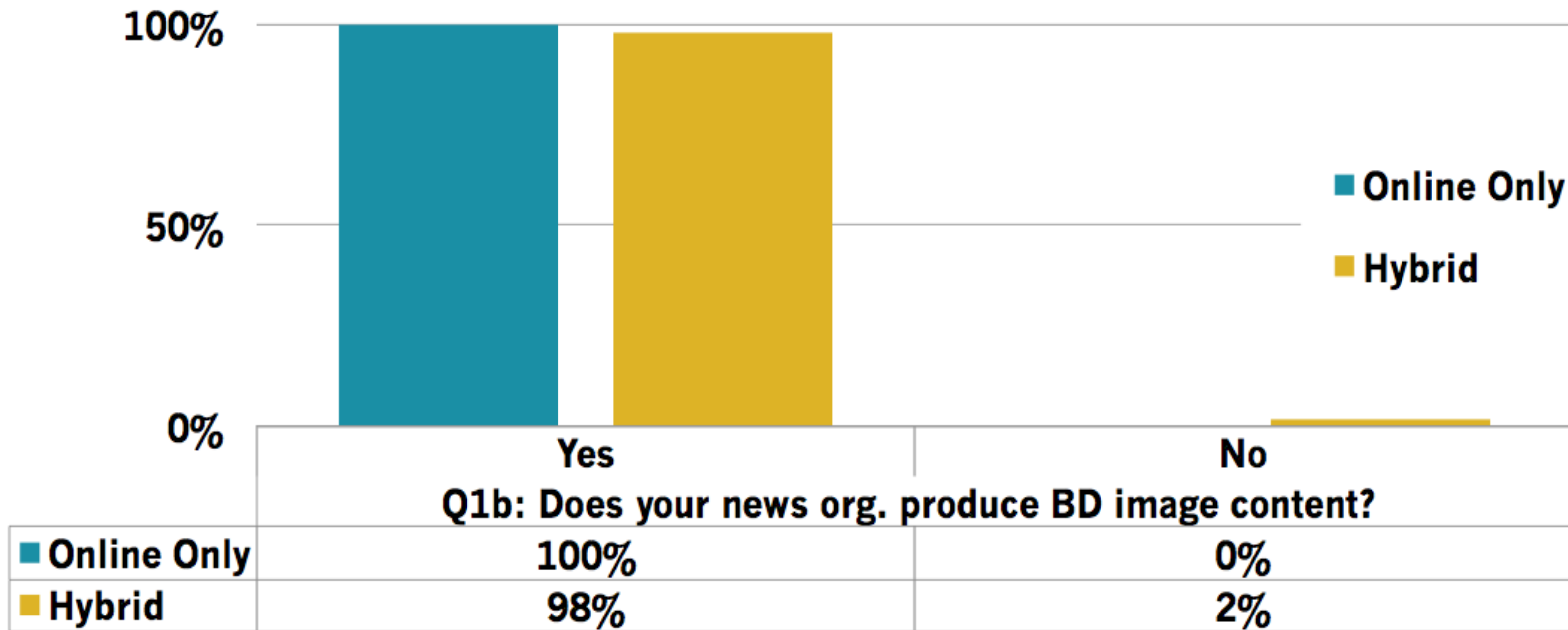
# Born-digital news text produced



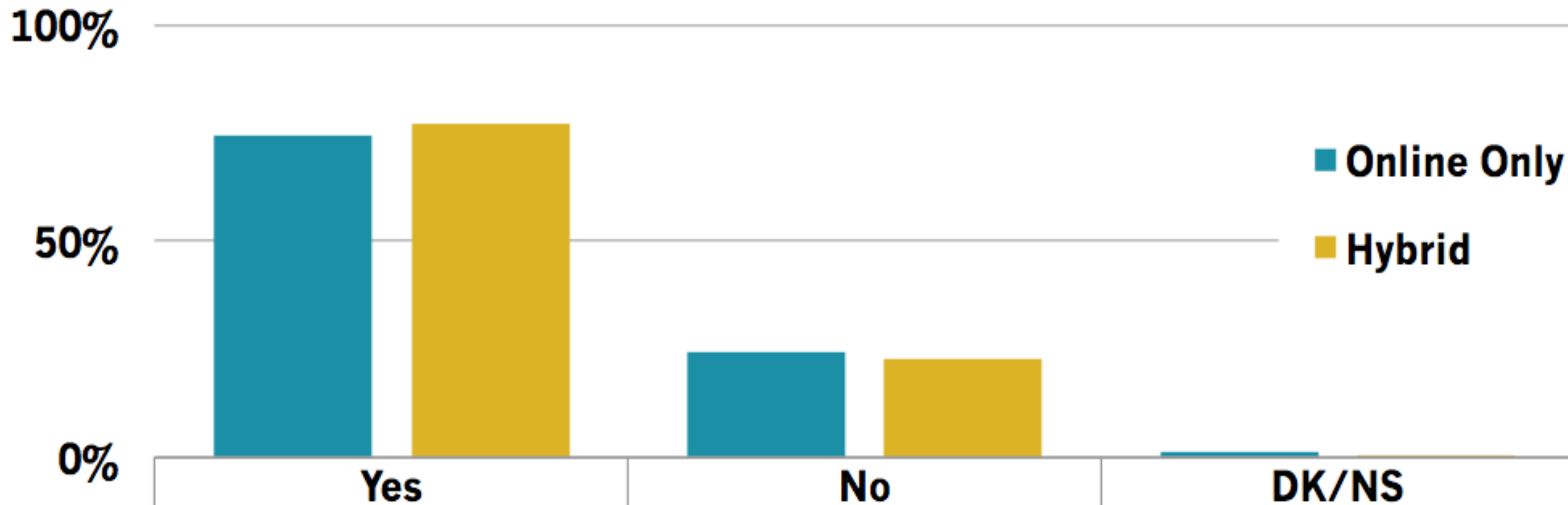
Q1a: Does your news org. produce BD text content?

■ Online Only	99%	1%	0%
■ Hybrid	93%	6%	1%

# Born-digital news images produced



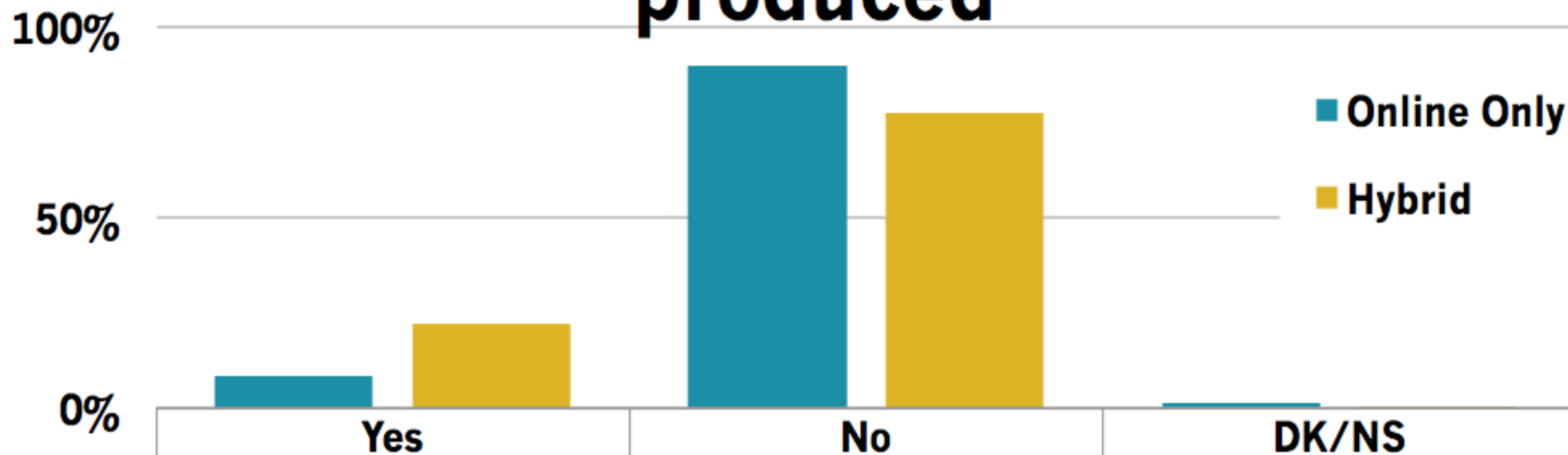
# Born-digital news video produced



**Q1c: Does your news org. produce BD video content?**

■ Online Only	74%	24%	1%
■ Hybrid	77%	23%	0%

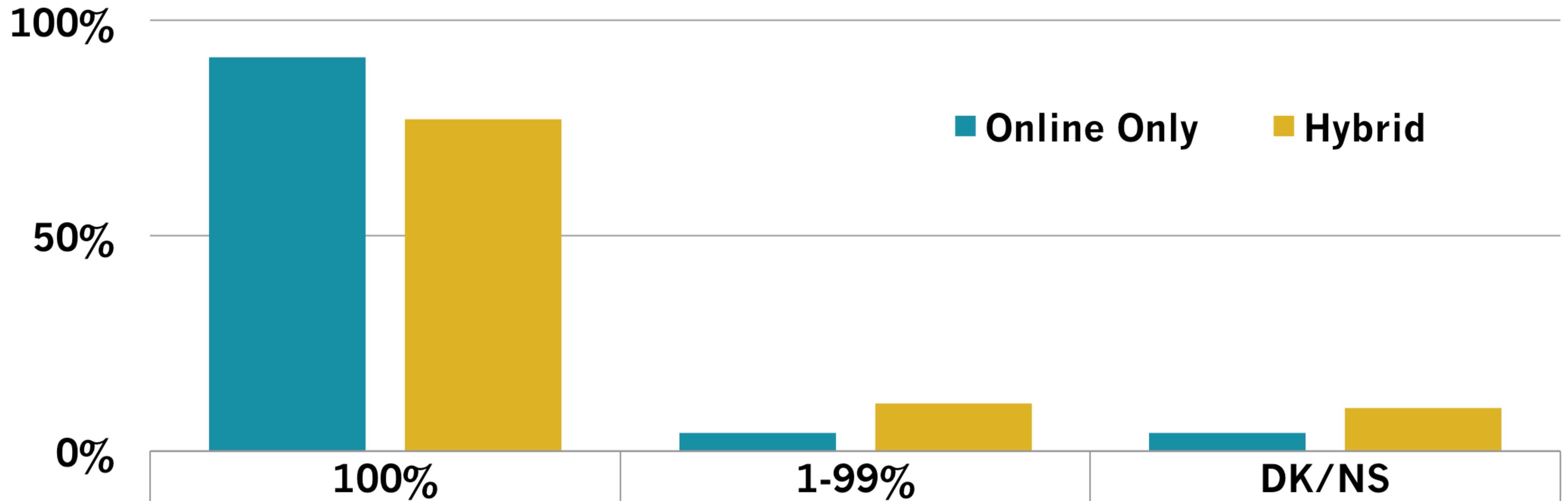
# Born-digital news mobile-only produced



**Q1e: Does your news org. produce mobile-only content?**

■ Online Only	9%	90%	1%
■ Hybrid	22%	77%	0%

# Amount of BDNC Backed Up

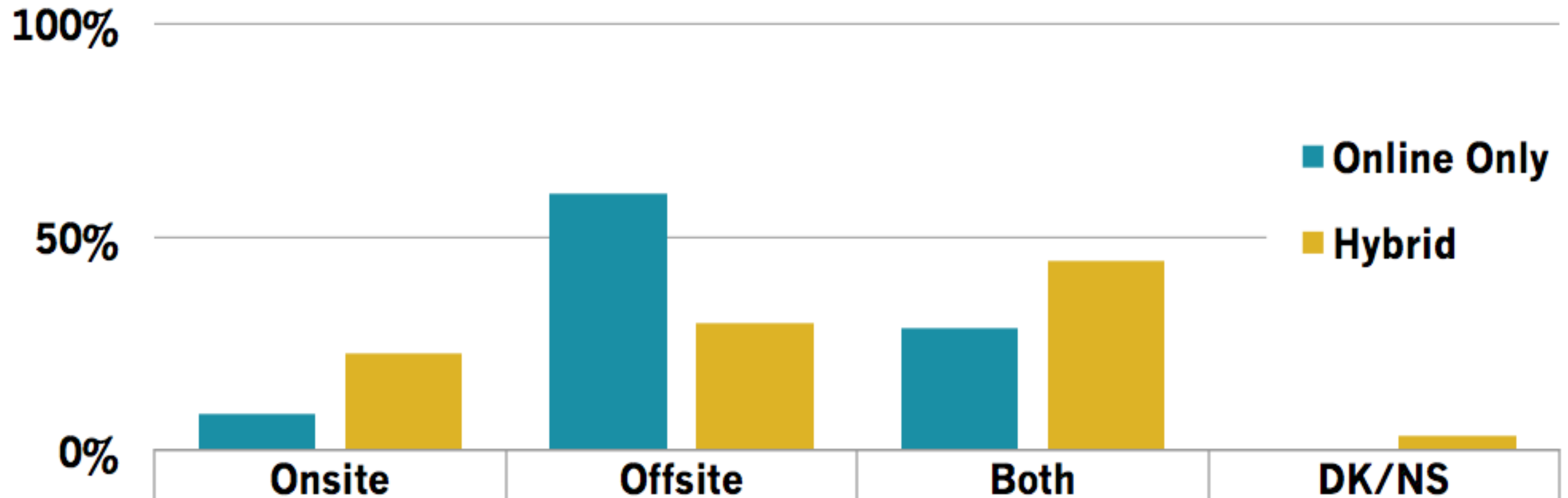


**100%**                      **1-99%**                      **DK/NS**

**Q2: How much of your BD news content is backed up?**

<b>■ Online Only</b>	<b>91%</b>	<b>4%</b>	<b>4%</b>
<b>■ Hybrid</b>	<b>77%</b>	<b>11%</b>	<b>10%</b>

# BDNC backed up on-site or off-site

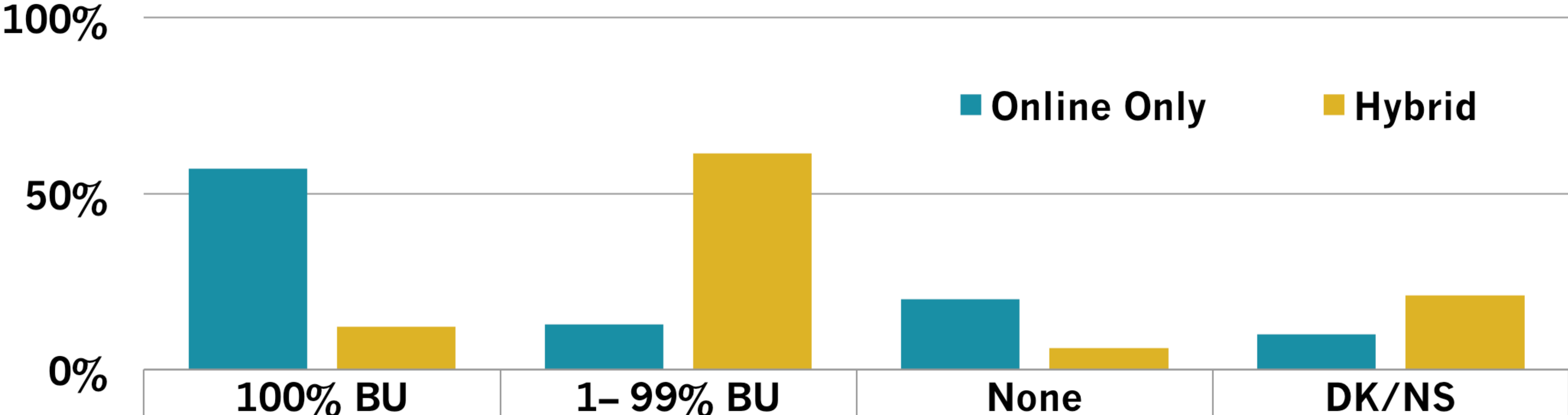


Q5e: Is the data backed up on- or off-site?

■ Online Only	9%	60%	29%	0%
■ Hybrid	23%	30%	44%	3%



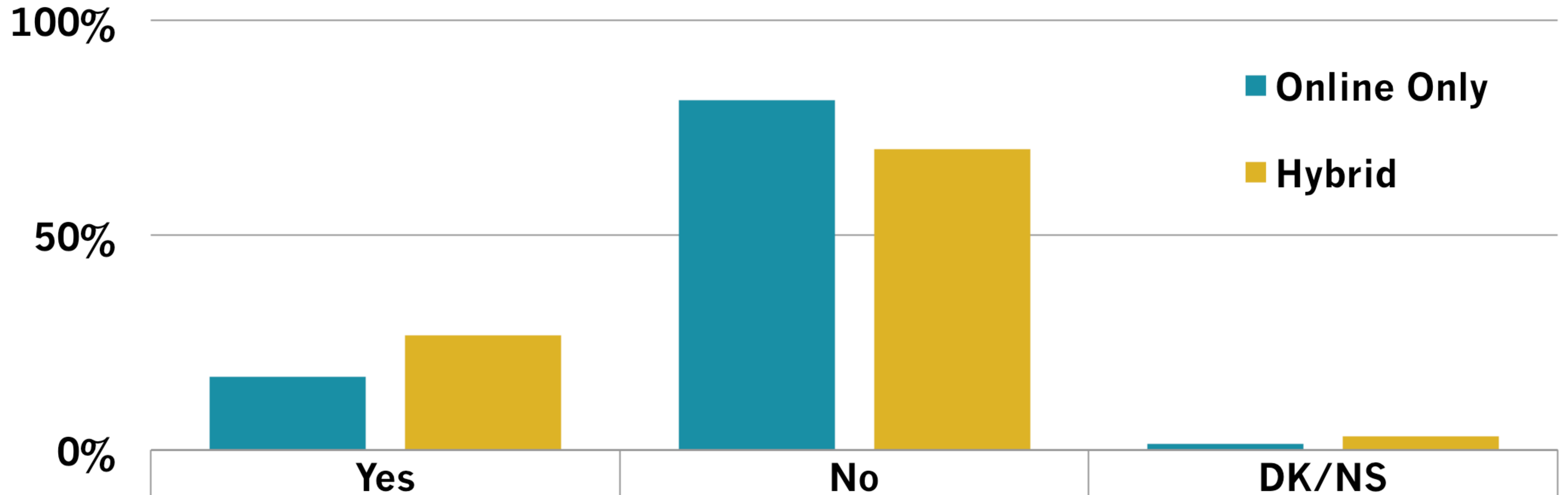
# What % BDNC Backed Up Last 25 Years



Q5b: About what % of your BDNC from the past 25 years is backed up?

■ Online Only	57%	13%	20%	10%
■ Hybrid	12%	61%	6%	21%

# Significant Loss of News Content



**Q5c: Have you ever experienced a significant loss of news content?**

■ Online Only	17%	81%	1%
■ Hybrid	27%	70%	3%

# Org. digital preservation program

100%

50%

0%

■ Online Only

■ Hybrid

Yes

No

DK/NS

Q6c: Does your org. have a digital preservation program?

■ Online Only

6%

33%

7%

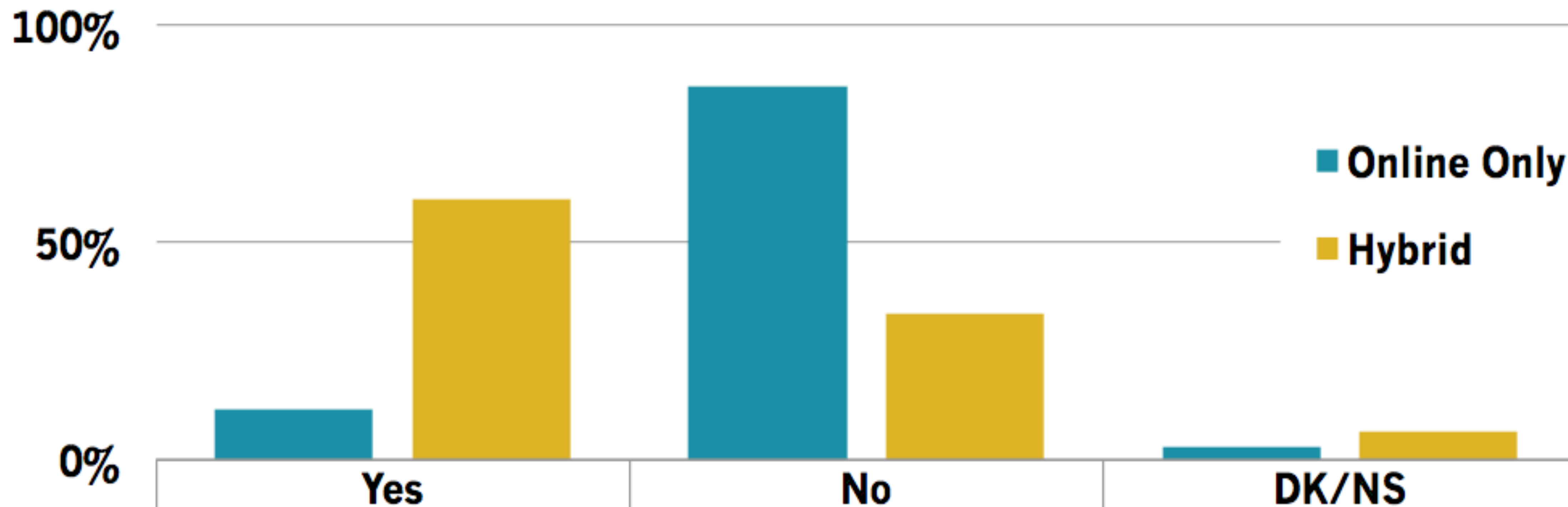
■ Hybrid

15%

34%

21%

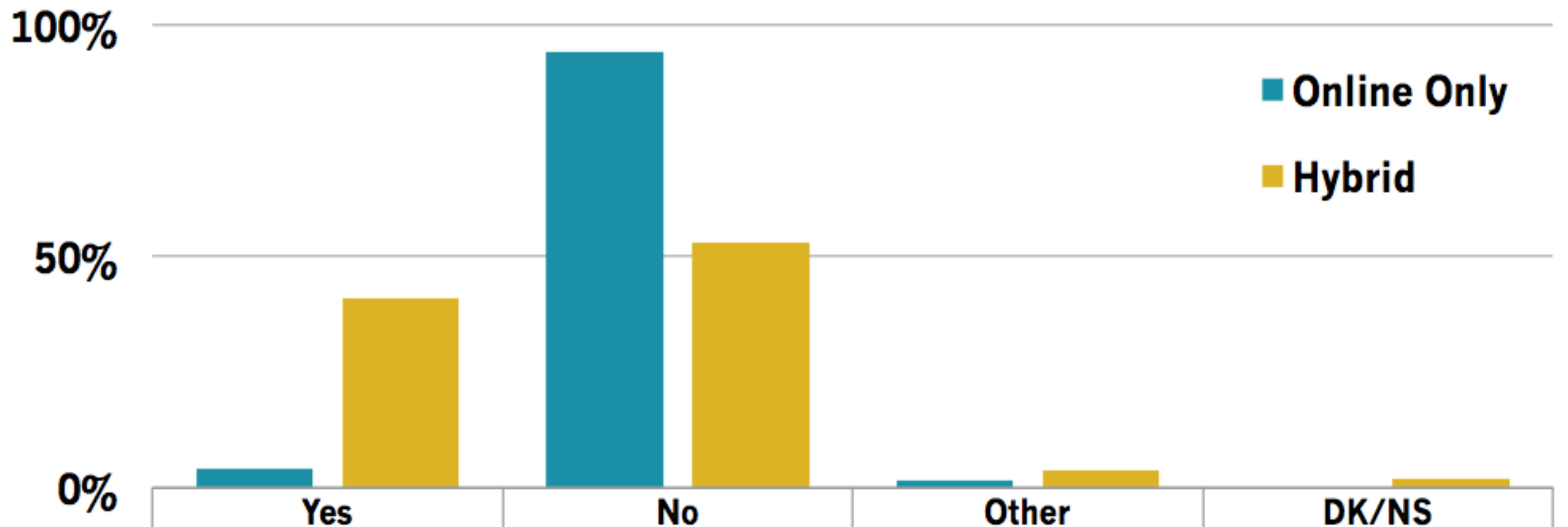
# BDNC going to memory institutions



Q6: Does your content go to a memory institution?

■ Online Only	11%	86%	3%
■ Hybrid	60%	34%	6%

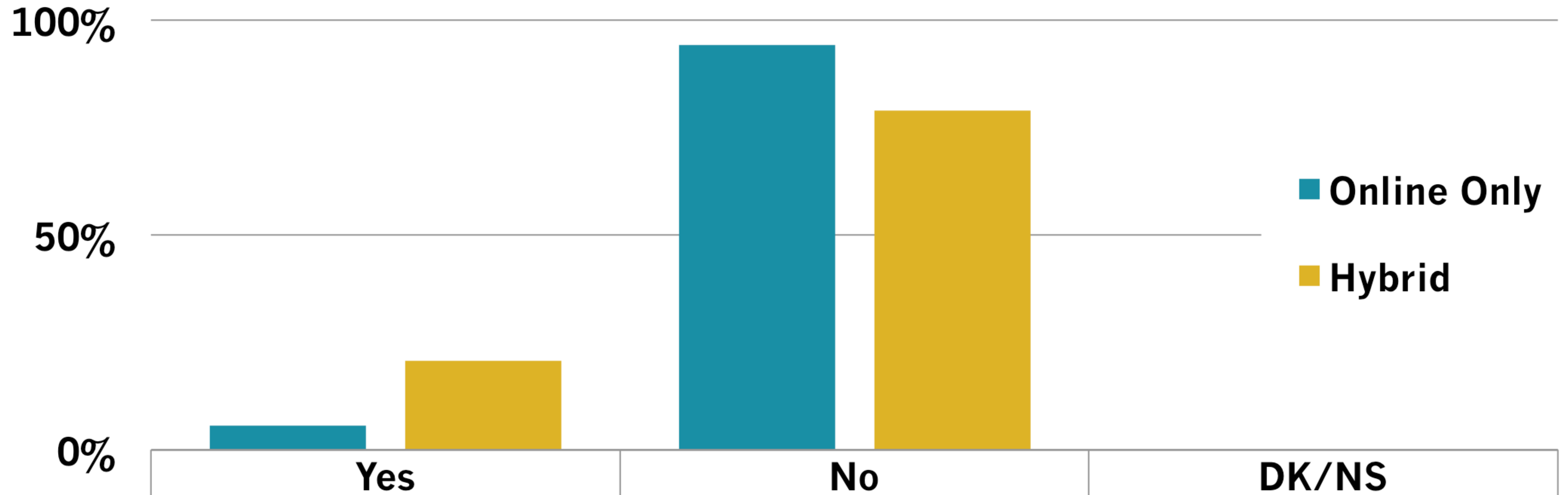
# Community news behind paywall



Q15a: Do you place obituaries, birth notices and marriage announcements behind a paywall?

■ Online Only	4%	94%	1%	0%
■ Hybrid	41%	53%	4%	2%

# Does org. have a news librarian?

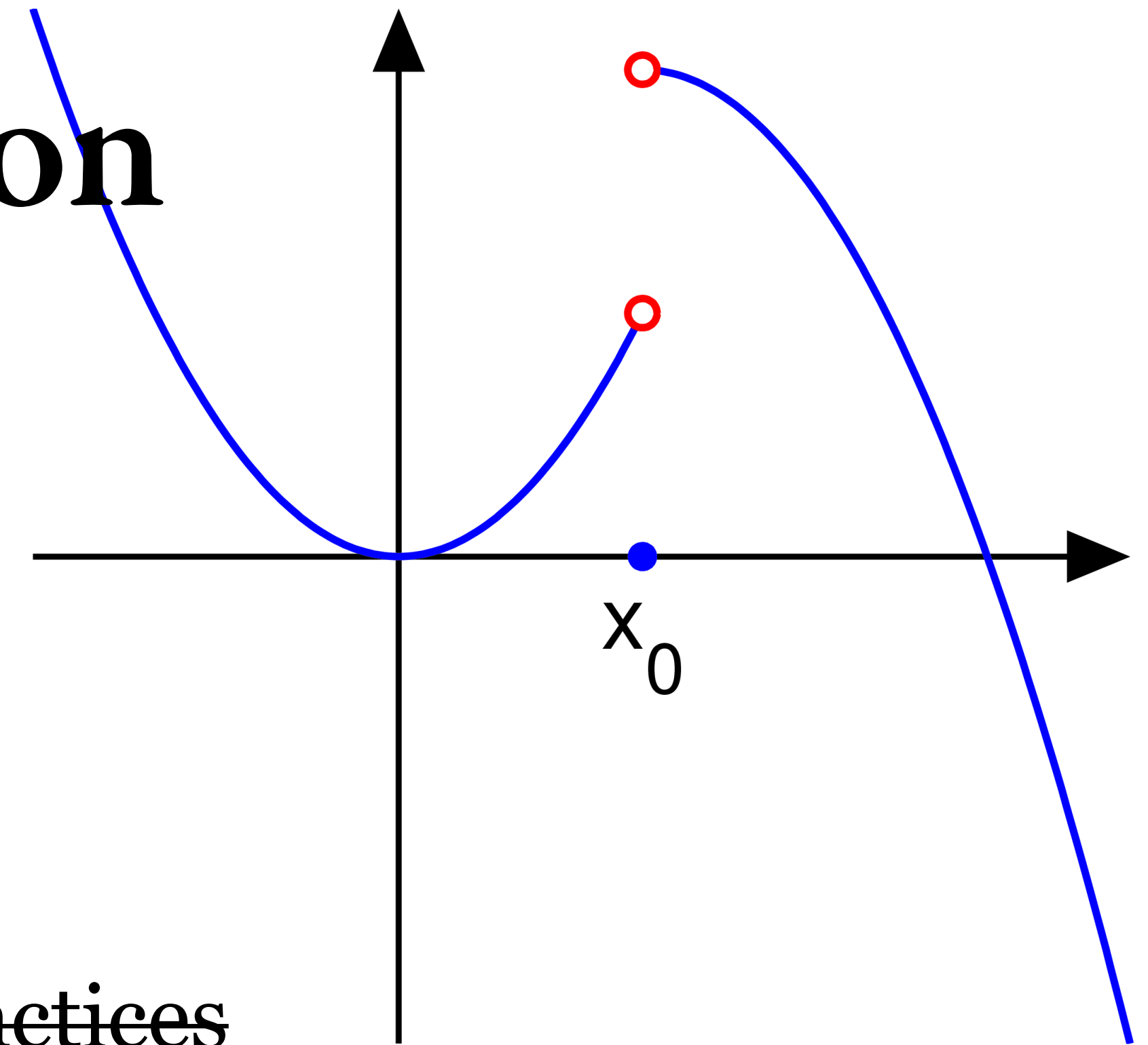


Q15c: Do you have a news librarian or equivalent position in your newsroom?

■ Online Only	6%	94%	0%
■ Hybrid	21%	79%	0%

# The digital news preservation

## DISCONTINUITY



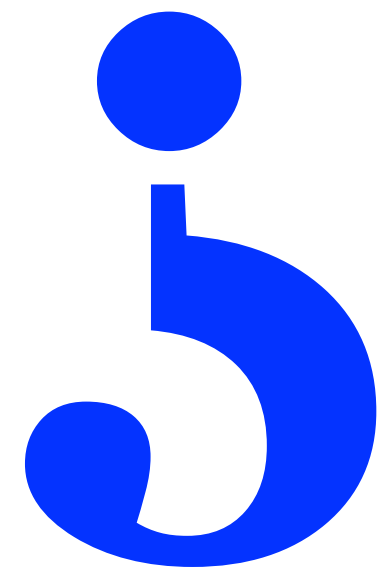
In this study we survey ...

- ~~... historical news preservation policies and practices~~
- ~~... born digital news preservation policies and practices of publishers~~
- **... cultural heritage organizations around the world about born digital news preservation policies and practices in their country**

# Questions for digital librarians

- A question for digital curators: What online born digital news should be preserved? (Remember who uses it.)
- A question for web harvesters: How many times per hour / day / week / month / year must a news website be harvested to capture X% of the news published on it?





# Questions for a future survey of news publishers



- How many versions of a news story does your organization typically produce?
- How long is an online news story live, that is, how long is it accessible to users?
- Does your news organization support **permalinks**?

According to Wikipedia a **permlink** is “a URL that points to a specific web page, often a blog or forum entry which has passed from the front page to the archives, or the result of a search in a database. Because a permlink remains unchanged indefinitely, it is less susceptible to link rot.”

questions?

вопросы?

问题吗?

Fragen?

¿preguntas?

الأسئلة؟

questions?

Dorothy Carner  
University of Missouri-Columbia  
[carnerd@missouri.edu](mailto:carnerd@missouri.edu)

Edward McCain  
Donald W. Reynolds Journalism Institute  
University of Missouri-Columbia  
[mccaine@rjionline.org](mailto:mccaine@rjionline.org)

Frederick Zarndt  
IFLA Newspapers Section  
[frederick@frederickzarndt.com](mailto:frederick@frederickzarndt.com)