

LA-UR-20-29413

Approved for public release; distribution is unlimited.

Title: Summarizing web archives through storytelling with the Dark and Stormy Archives Project

Author(s): Jones, Shawn Morgan

Intended for: Coalition for Networked Information Fall 2020 Meeting,  
2020-11-10/2020-11-13 (Washington, District Of Columbia, United States)

Issued: 2020-11-17

---

**Disclaimer:**

Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by Triad National Security, LLC for the National Nuclear Security Administration of U.S. Department of Energy under contract 89233218CNA000001. By approving this article, the publisher recognizes that the U.S. Government retains nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.

# Summarizing web archives through storytelling with the Dark and Stormy Archives Project

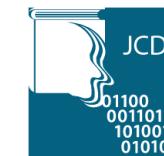
Shawn M. Jones

Los Alamos National Laboratory  
Research Library Prototyping Team

Web Science and Digital Libraries Research Group  
Old Dominion University

CNI Fall 2020 Meeting

Thanks to:

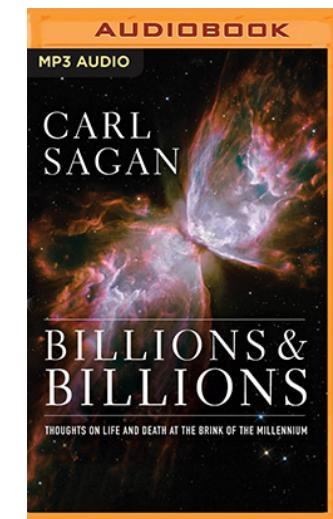




# The abundance of data in web archives



@shawnmjones @StormyArchives



# Web archives contain numerous documents



Brewster Kahle  
@brewster\_kahle

...

Wayback Machine now has 898,570,440,000 URL's -- serving millions of users as a foundation for what has been said on a constantly shifting WWW. (up 17 billion URLs in less than 1 month ago)

[twitter.com/brewster\\_kahle...](https://twitter.com/brewster_kahle/status/1225167435399036939) ) Go [@internetarchive](https://twitter.com/brewster_kahle) !



Brewster Kahle @brewster\_kahle · Jan 10

Wayback Machine just grew to 881,352,519,000 web URL's. That is 881 Billion. For every one that becomes important in the news or in someones personal world, we crawl and store millions of others just-in-case. go [@internetarchive](https://twitter.com/brewster_kahle)

2:20 PM · Feb 5, 2020 · Twitter Web App

- [https://twitter.com/brewster\\_kahle/status/1225167435399036939](https://twitter.com/brewster_kahle/status/1225167435399036939)
- <https://www.webarchive.org.uk/en/ukwa/info/faq>
- <https://archive-it.org/collections/3572>

## How big is the archive?

The UK Web Archive collects millions of websites each year and billions of individual assets (pages, images, videos, pdfs etc.). As of 2017 we have collected approximately 500TB of data and increasing this by over roughly 60 – 70 TB a year.



### Government of Canada Publications

Collected by: Canadian Government Information PLN Web Archive

Archived since: Feb, 2013  
No description.

#### Narrow Your Results

Creator Sort By: Count | (A-Z)

- Fisheries and Oceans Canada (867)
- House of Commons (627)
- Environnement Canada (456)
- Environment Canada (444)
- Chambre des communes (443)

More ▾

Sites for this collection are listed below. Narrow your results at left, or enter a search query below to find a site, specific URL or to search the text of archived webpages.

Enter search terms here

Search

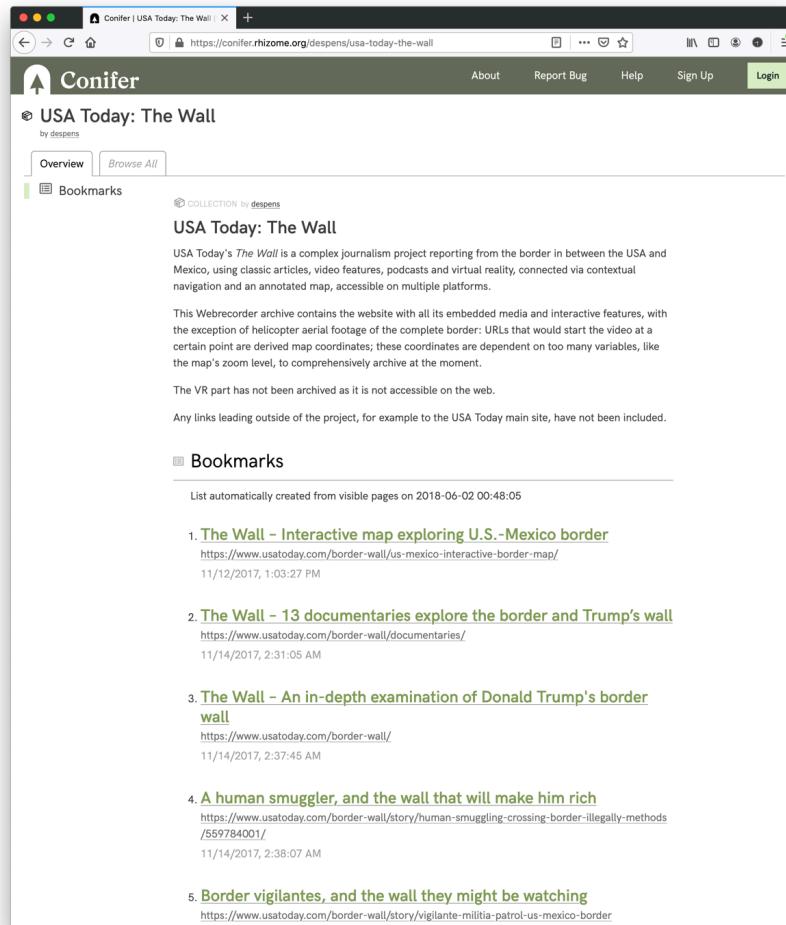
Clear

# 339,166 Total Results



@shawnmjones @StormyArchives

# Mementos are the documents in web archive collections



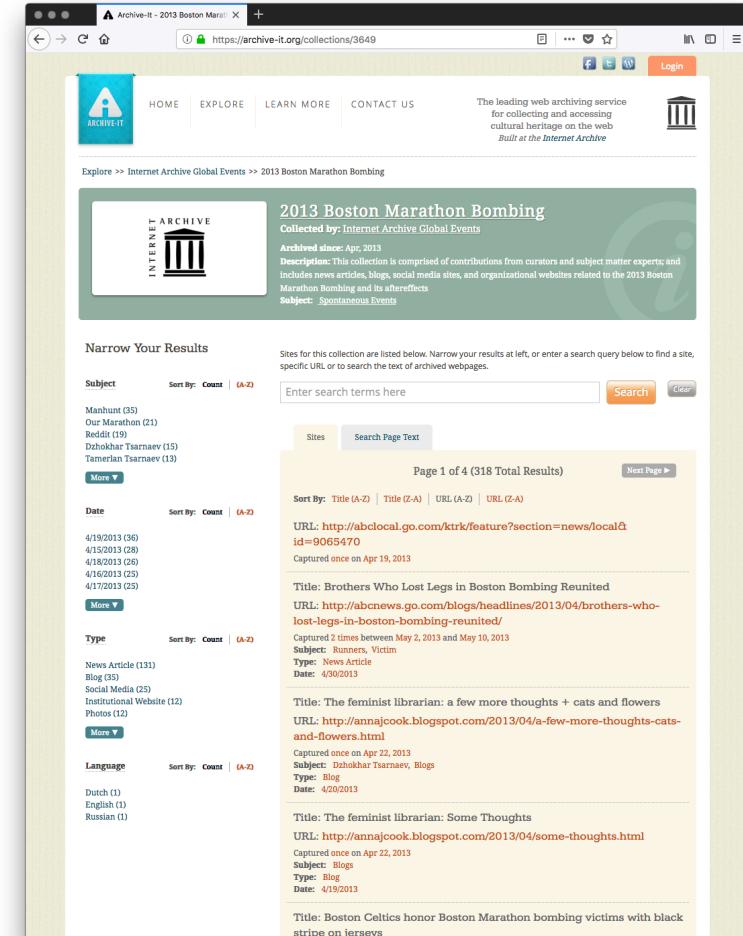
The screenshot shows a web browser window with the title 'Conifer | USA Today: The Wall'. The page displays a collection of web pages under the heading 'USA Today: The Wall'. It includes sections for 'Overview', 'Bookmarks', and a list of 5 items. Item 1: 'The Wall - Interactive map exploring U.S.-Mexico border' (https://www.usatoday.com/border-wall/us-mexico-interactive-border-map/). Item 2: 'The Wall - 13 documentaries explore the border and Trump's wall' (https://www.usatoday.com/border-wall/documentaries/). Item 3: 'The Wall - An in-depth examination of Donald Trump's border wall' (https://www.usatoday.com/border-wall/). Item 4: 'A human smuggler, and the wall that will make him rich' (https://www.usatoday.com/border-wall/story/human-smuggling-crossing-border-illegally-methods/559784001/). Item 5: 'Border vigilantes, and the wall they might be watching' (https://www.usatoday.com/border-wall/story/vigilante-militia-patrol-us-mexico-border/559784001/).

Collections with Conifer/Webrecorder

Web archive collections consist of **seeds** and crawls of those seeds.

The **mementos** from these crawls are versions of pages from the time of the crawl.

The mementos are the documents in our collections.



The screenshot shows a web browser window with the title 'Archive-It - 2013 Boston Marathon Bombing'. The page displays a collection of web pages under the heading '2013 Boston Marathon Bombing'. It includes sections for 'Explore', 'Archive-It ARCHIVE', and a list of results. The results table has columns for 'Subject', 'Date', 'Type', and 'Language'. The results list includes items such as 'Brothers Who Lost Legs in Boston Bombing Reunited' (Captured on Apr 19, 2013), 'The feminist librarian: a few more thoughts + cats and flowers' (Captured on Apr 22, 2013), and 'Boston Celtics honor Boston Marathon bombing victims with black stripe on jerseys' (Captured on Apr 22, 2013).

Collections with Archive-It



@shawnmjones @StormyArchives

# Web archives have many versions (mementos) of the same page and *this is a feature...*



2013



2015

2018

University of Utah Office of Admissions  
from the University of Utah Web Archive Collection



4/1/2015

Tumblr Black Lives Matter Blog  
from the #blacklivesmatter Collection



@shawnmjones @StormyArchives

**Different versions (mementos) allow us to see an unfolding news story**

Boston Marathon Suspect De... +

wayback.archive-it.org/3649/20130419171223/http://bostinno.streetwise.co/

**Bostinno**

**GET THE BOSTINNO NEWSLETTER**  
Email Address...

**COMMUNITIES** **CHANNELS** **EVENTS** **SERIES** **CAREERS** [Log In](#) [Signup](#)

**TRENDING** • Resources Post-Marathon... • Yankees Fans Sing Sweet... • MIT's Green Building is Lit... • Two Bombs Explode at the... • Nearly 100 Greater Boston...

**FRONTPAGE**  
**ARTICLES**  
**MEMBER CHANNELS**  
**CHANNELS**

**LOCAL COMMUNITY**

## One Boston Marathon Suspect Killed, Another At Large [Live Updates]

Lauren Landry *Today at 7:03 am*  
News, *MIT Boston Marathon 2013*



**LATEST ACTIVITY**

**Carla Pichard** gave +1 to an article  
**Join Us: A Special Meet Up to Benefit Marathon Attack Victims...**

**Robert McElney** gave +1 to an article  
**President of Chechnya Criticizes Boston...**

**John Whittle** gave +1 to an article  
**Join Us: A Special Meet Up to Benefit Marathon Attack Victims...**

**Robert McElney** gave +1 to an article  
**President of Chechnya Criticizes Boston...**

**John Whittle** gave +1 to an article  
**Join Us: A Special Meet Up to Benefit Marathon Attack Victims...**

Alben followed by reporting authorities are going door to door, street to street in Watertown. A few new leads have emerged, which will be made aware to the public in roughly one hour. Alben also announced there will, indeed, be a controlled explosion on Norfolk Street in Cambridge this afternoon in a house they secured earlier today.

100+ more great stories delivered to the Watertown Community. You've done everything we've asked, but we need some more time."

Deveau assured the town authorities will not be leaving until the neighborhood is safe, and Menino said they will keep the public informed as news comes in.

Robert McElney gave +1 to an article  
**Suspect Photos Released in Boston Marathon**

# Memento from April 19, 2013 17:12

## Searching for suspects, City on lockdown

Boston Marathon Suspect Dead

wayback.archive-it.org/3649/20130419175943/http://bostinno.streetwise.co...

**BostInno**

SEARCH

LOG IN | SIGN UP

GET THE BOSTINNO NEWSLETTER

EMAIL ADDRESS...

SUBSCRIBE

**COMMUNITIES** | **CHANNELS** | **EVENTS** | **SERIES** | **CAREERS**

**TRENDING** • Resources Post-Marathon... • Yankees Fans Sing Sweet... • MIT's Green Building is Lit... • Two Bombs Explode at the... • Nearly 100 Greater Boston...

FRONT PAGE | ARTICLES | MEMBER CHANNELS | CHANNELS

LOCAL COMMUNITY

## One Boston Marathon Suspect Killed, Another At Large [Live Updates]

Lauren Landry Today at 7:09 am  
News, MIT, Boston Marathon 2013

**Update, 1:43 p.m.** Wounded Transit Police Officer Richard H. Donohue Jr. has been upgraded to stable condition, according to 7News, who report he was visited by Governor Patrick today around 1 p.m.



**Update, 1:40 p.m.** Arkansas Police Officer Thinks 'Cowering' Bostonians half apologize around the 'liming' of the statement several hours later.

**Update, 1:40 p.m.** Kurt Schwartz, director of the Massachusetts Emergency Management Agency, made an announcement to the media, saying that all those who found their way to work before the lockdown went into effect are able to leave and go home. Although the MBTA is still not in service, cabs are running in the city of Boston.

Update, 1:40 p.m. Kurt Schwartz, director of the Massachusetts Emergency Management Agency, made an announcement to the media, saying that all those who found their way to work before the lockdown went into effect are able to leave and go home. Although the MBTA is still not in service, cabs are running in the city of Boston.

**Boston Red Sox** players and other team employees have been told to stay home until a decision is made on whether tonight's game will be played, according to the *Associated Press*. The night game between the Boston Bruins and Pittsburgh Penguins is also on hold after the NHL teams canceled their

**Elizabeth Weaver** commented on Suspect Photos Released in Boston Marathon

**John** commented on How to Help Victims & Authorities in the Wake of Boston's...

# Memento from April 19, 2013 17:59

Officer Donahue in hospital,  
Lockdown loosened,  
Will the Red Sox game be cancelled?

**Boston Marathon Suspect Dies**

wayback.archive-it.org/3649/20130421022426/http://bostinno.streetwise.co/

**BostInno**

COMMUNITIES CHANNELS EVENTS SERIES CAREERS

FRONT PAGE ARTICLES MEMBER CHANNELS CHANNELS

LOCAL COMMUNITY

## One Boston Marathon Suspect Killed, Another At Large [Live Updates]

Lauren Landry Today at 7:09 am  
News, MIT, Boston Marathon 2013

**Updated, 10:20 p.m.** "We have closed an important chapter in this tragedy," said President Barack Obama in a press conference at the White House, minutes after a briefing came to a close in Watertown.

Obama acknowledged the coordination among state, federal and local officials, proclaiming, "They all worked as they should—as a team. We are extremely grateful for that."

He also honored Sean Collier, the MIT police officer who lost his life at the hand of the Marathon bombing suspects, and acknowledged the questions that remain.

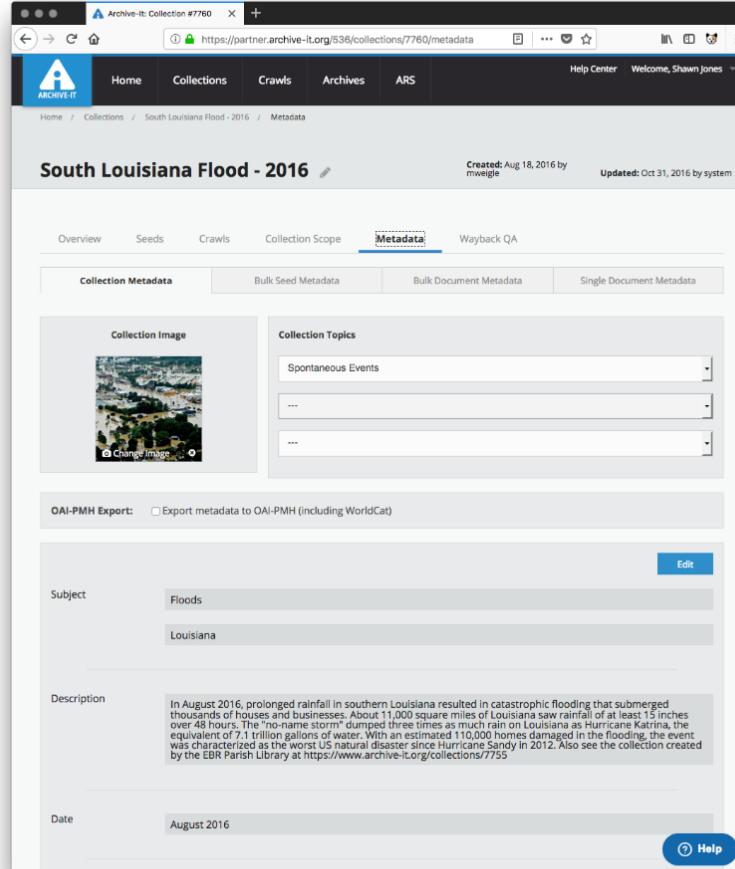
*Why did young men who grew up and study here...resort to such violence?*

*How did they plan and carry out these attacks, and did they receive any help?*

"The wounded deserve answers," Obama said, announcing all necessary resources will continue to be deployed to determine what happened.

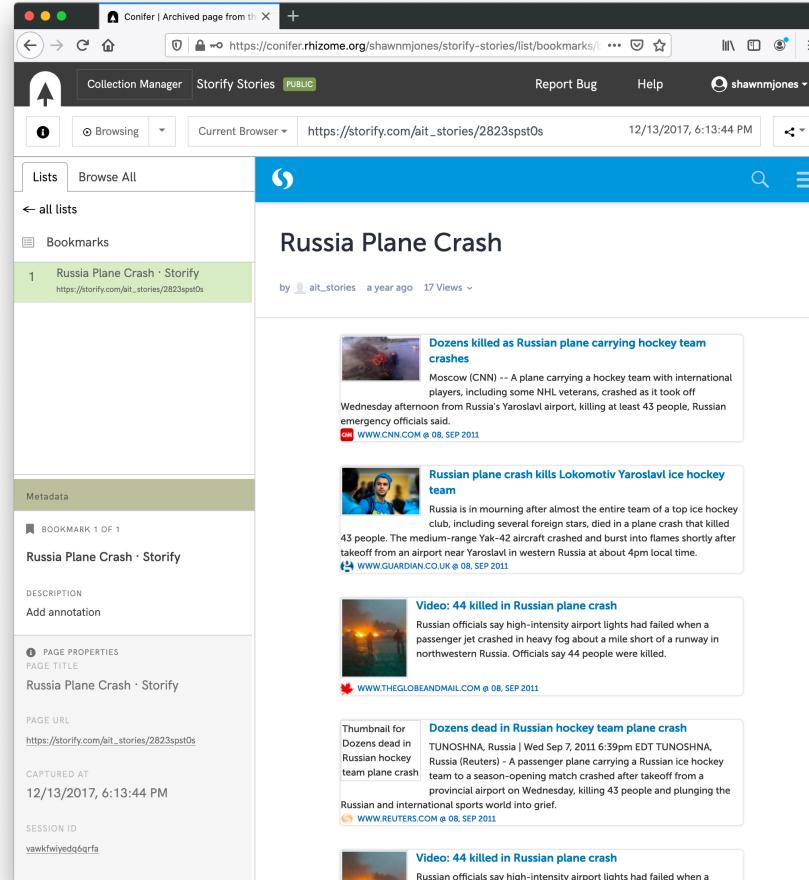
# Memento from April 24, 2013 2:24 Suspect Found, Officer Collier lost life, Obama speaks

# Tools exist that easily allow users to add content to web archives...



The screenshot shows the Archive-It interface for a collection titled "South Louisiana Flood - 2016". The top navigation bar includes links for Home, Collections, Crawls, Archives, ARS, Help Center, and a user account. The main content area displays the collection's metadata, including an "Overview" tab, "Collection Topics" (Spontaneous Events), and a "Description" section detailing the flooding in August 2016. The "Date" section shows the collection was created in August 2016.

Archive-It



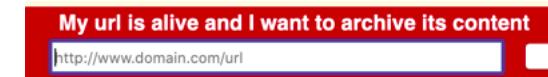
The screenshot shows the Conifer/Webrecorder interface. It displays a bookmarked page titled "Russia Plane Crash" from Storify. The page includes a thumbnail image, a summary, and links to news articles from CNN and The Guardian. The interface also shows the original page URL, capture date (12/13/2017, 6:13:44 PM), and session ID.

Conifer/Webrecorder



The screenshot shows the "Save Page Now" interface. It features a text input field for the URL, a "SAVE PAGE" button, and a descriptive text: "Capture a web page as it appears now for use as a trusted citation in the future. Only available for sites that allow crawlers."

## Internet Archive's Save Page Now



The screenshot shows the "Archive.Today's On-demand Web Form" interface. It has a red header with the text "My url is alive and I want to archive its content" and a "save" button. Below is a text input field with the URL "http://www.domain.com/url".

## Archive.Today's On-demand Web Form



# ... and these collections are used by other researchers

These collections live a life after their curator has stopped adding to them.

The collection curator is not the only user of the collection!

# Consider Archive-It as an example...

# There are many collections on the same topic: 31 Archive-It collections match the search query “human rights”

Which one is best for my needs?

Human Rights Documentation Initiative  
Collected by: University of Texas at Austin Libraries, Human Rights Documentation Initiative  
Archived since: Apr, 2009  
Description: The University of Texas Libraries' Human Rights Documentation Initiative Collection features fragile websites containing human rights documentation and related content from human rights organizations and advocates across the globe. Browse collection by topic, geographic area, language or do a keyword search at [http://www.lib.utexas.edu/hrdi/hr\\_archive](http://www.lib.utexas.edu/hrdi/hr_archive).  
Subject: Society & Culture, Politics & Elections, Blogs & Social Media

Narrow Your Results

Group Sort By: Count | (A-Z)

- Association africaine de Défense des droits de l'Homme (ASADHO) (2)
- Cambodian Center for Human Rights (CCHR) (2)
- Change for Equality (2)
- Chinese Human Rights Defenders (2)
- East Turkistan Information Center (2)

More ▾

Subject Sort By: Count | (A-Z)

- advocacy, activism, and responses to persecution—activists (54)
- civil and political rights—women's rights (48)
- armed conflict and persecution—torture (34)

Enter search terms here  Search

Sites for this collection are listed below. Narrow your results at left, or enter a search query below to find a site, specific URL or to search the text of archived webpages.

Sites

Page 1 of 3 (267 Total Results) [Next Page ▶](#)

Sort By: Title (A-Z) | Title (Z-A) | URL (A-Z) | URL (Z-A)

Title: Adaleh Center for Human Rights Studies  
URL: <http://adaleh-center.org/news/>  
Description: In Arabic and English, ADALEH Center for Human Rights is a non-governmental, non-profit organization founded in September 2003, with a mission of enforcing human rights in Jordan and Arab world. Projects include youth capacitation, training medical staff in torture detection, and capacitating lawyers in the field of human rights. Site contains news and publications.

Human Rights  
Collected by: Columbia University Libraries  
Archived since: May, 2008  
Description: An initiative of CUL's Center for Human Rights Documentation and Research, the Human Rights Web Archive is an effort to preserve and ensure access to freely available human rights resources created mainly by non-governmental organizations, national human rights institutions, and individuals.  
Subject: Society & Culture, Human rights, Non-governmental organizations, Human rights workers, National human rights institutions, Web archives

Human Rights  
Collected by: University of Manitoba  
Archived since: Apr, 2014

Search

Search query below to find a site, or enter a search query below to find a site, the text of archived webpages.

Search

Page 1 of 1 (48 Total Results) [Next Page ▶](#)

Sort By: Title (A-Z) | Title (Z-A) | URL (A-Z) | URL (Z-A)

North Carolina  
ofnc.org/  
ween Dec 12, 2011 and Mar 3, 2016  
ured  
Civil Liberties Union of North Carolina  
antenc.org/

How are they different from each other?



@shawnmjones @StormyArchives

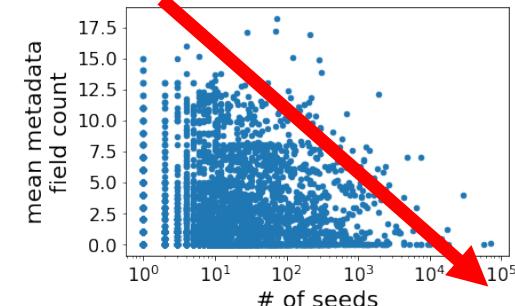
# Metadata is there, but does not always help

Because metadata is optional it is not always present.

Metadata on Archive-It collections:

- many different curators
- different organizations
- different content standards
- different rules of interpretation
- ***it is inconsistently applied***

***This means that a user cannot reliably compare metadata fields to understand the differences between collections.***



9 seeds  
with metadata

Paradox:  
More seeds = more effort  
More seeds = greater user need for metadata

132,599 seeds  
no metadata

Jones, S.M., Weigle, M.C., & Nelson, M.L. (2019). Social Cards Probably Provide For Better Understanding Of Web Archive Collections. In ACM CIKM '19. <https://doi.org/10.1145/3357384.3358039>



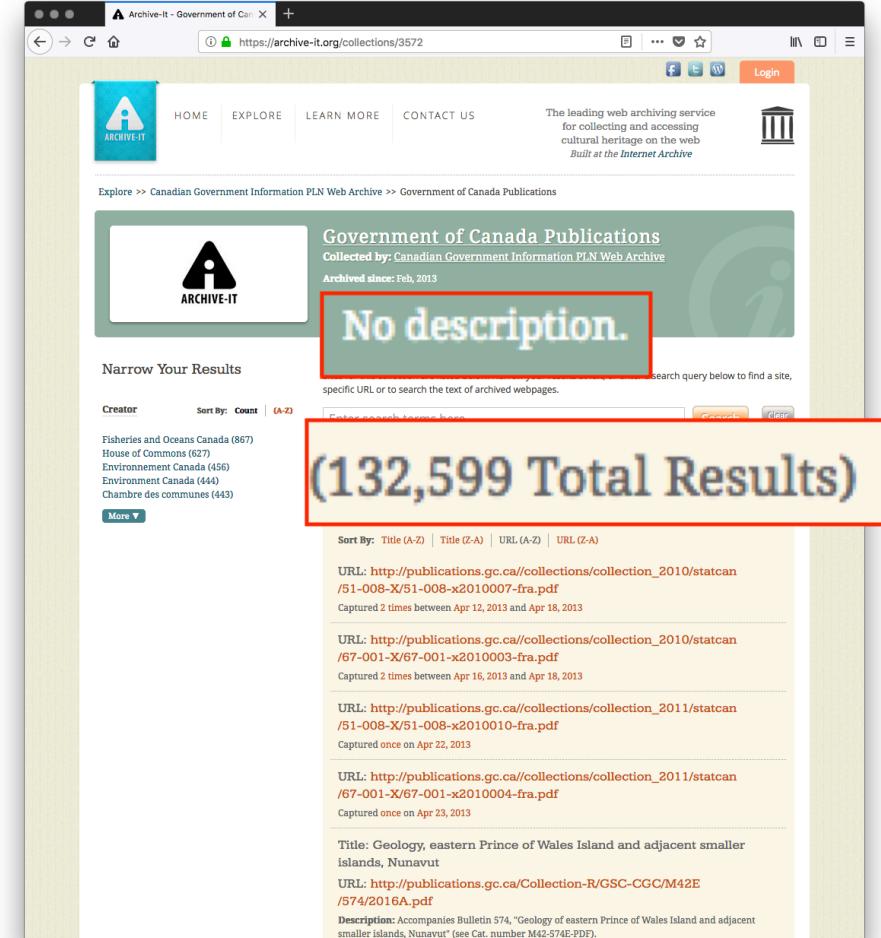
@shawnmjones @StormyArchives

# Reviewing mementos manually is costly

Some collections have 1000s of seeds

Each seed can have many mementos

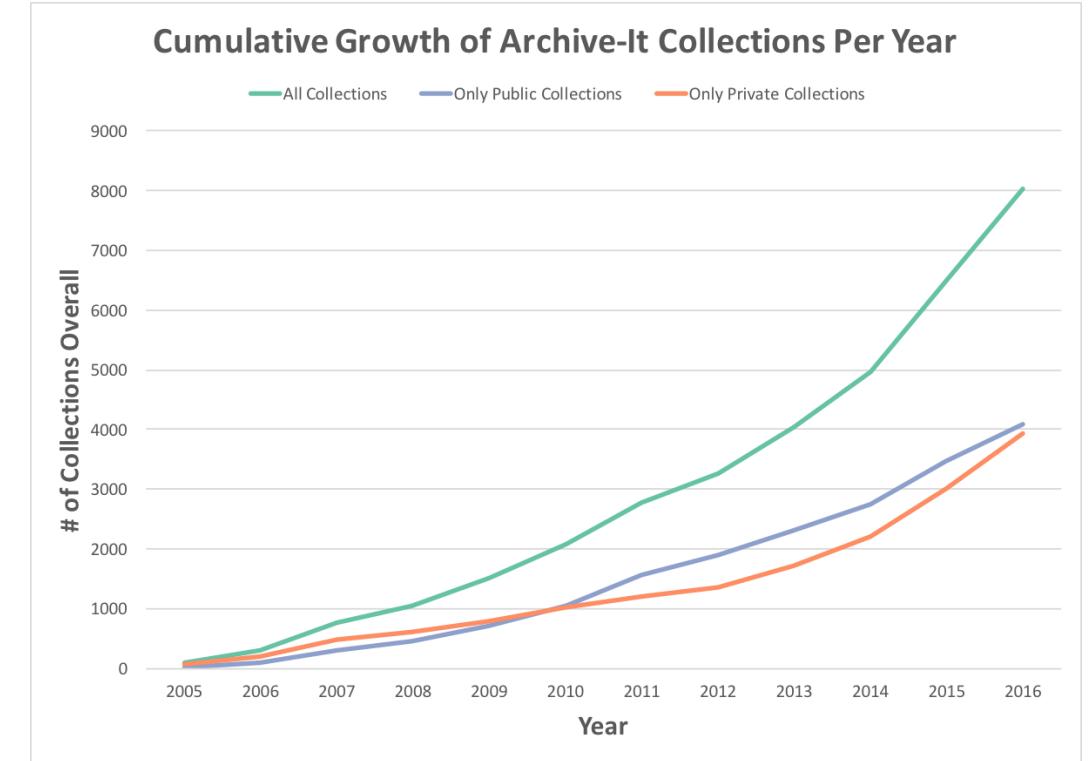
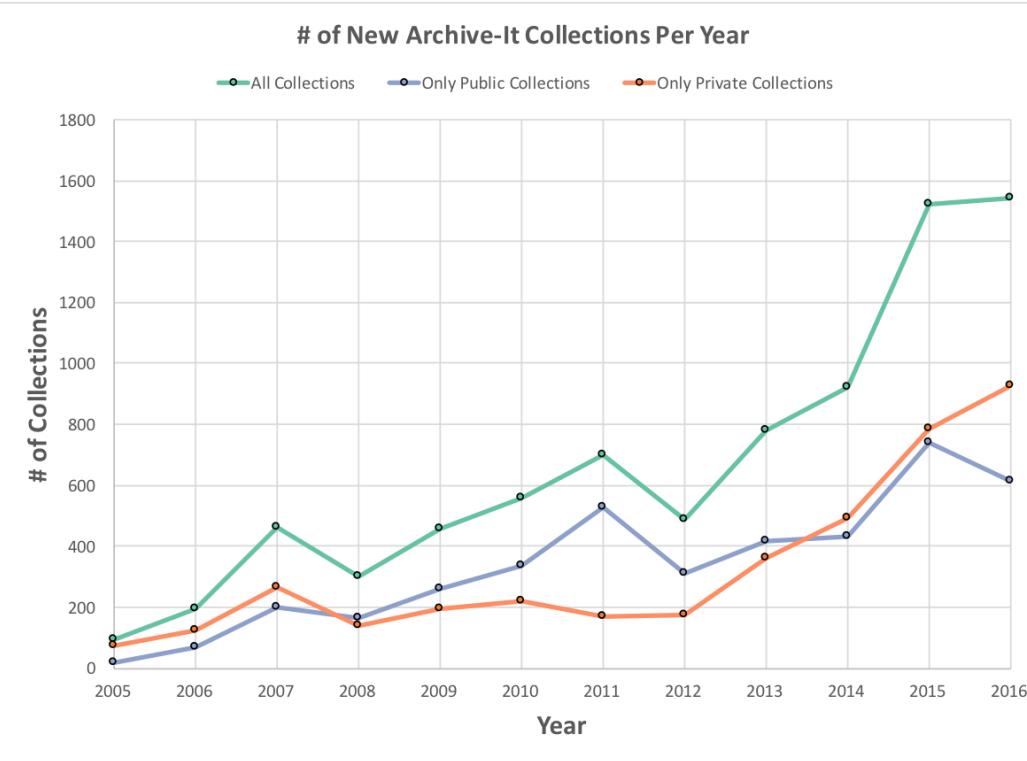
In some cases, this can require reviewing 100,000+ documents to understand the collection



This collection has 132,599 seeds, many with multiple mementos



# In addition to more mementos in web archives, more Archive-It collections are added every year

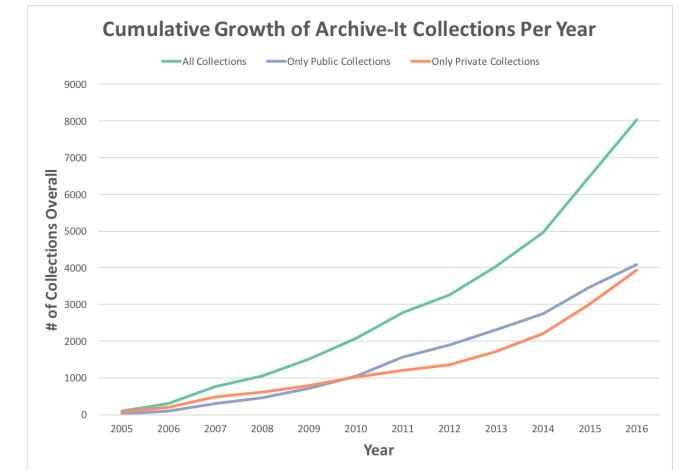
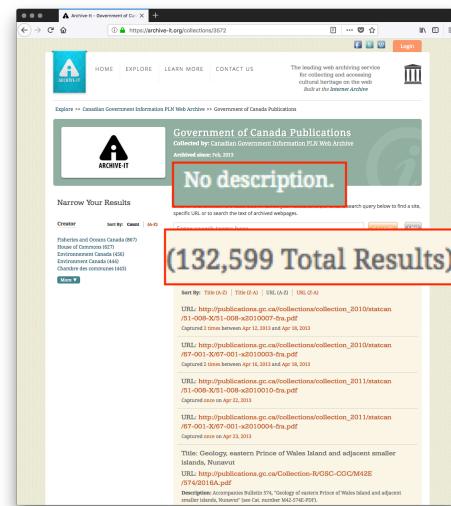
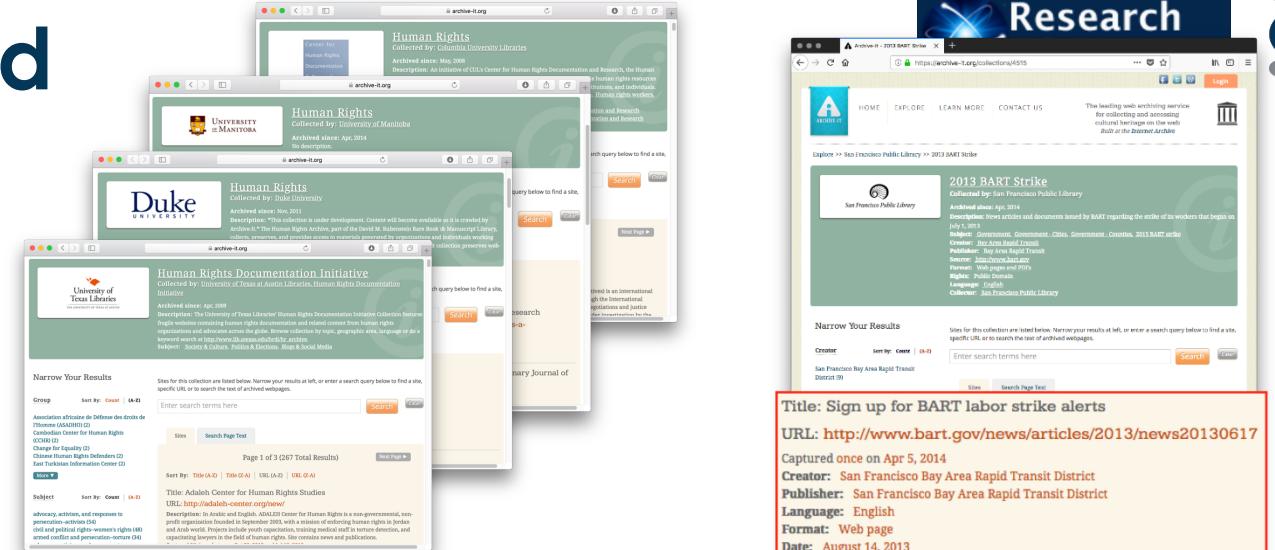


More than 15,000 collections exist



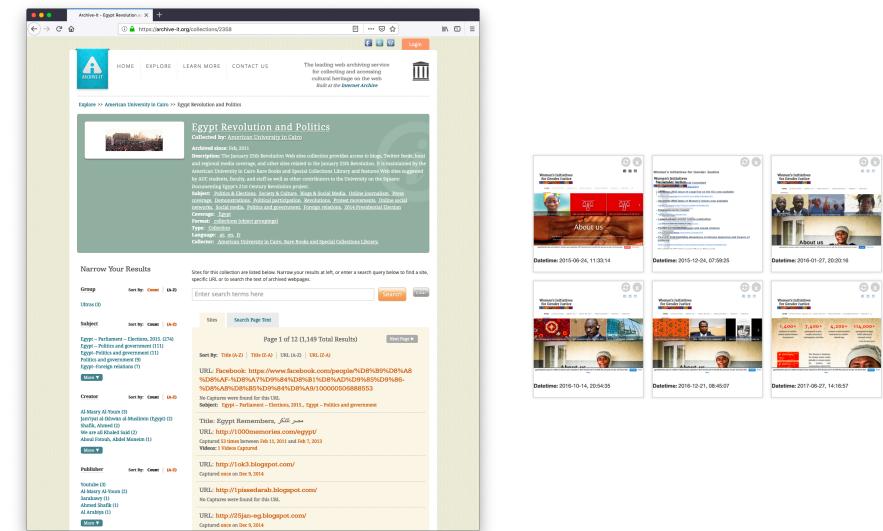
# The problem, summarized

- There are multiple collections about the “same concept.”
- The metadata for each collection is non-existent, or inconsistently applied.
- Many collections have 1000s of seeds with multiple mementos.
- There are more than 15,000 collections.
- ***Human review of these mementos for collection understanding is prohibitively expensive.***





# Existing Solutions

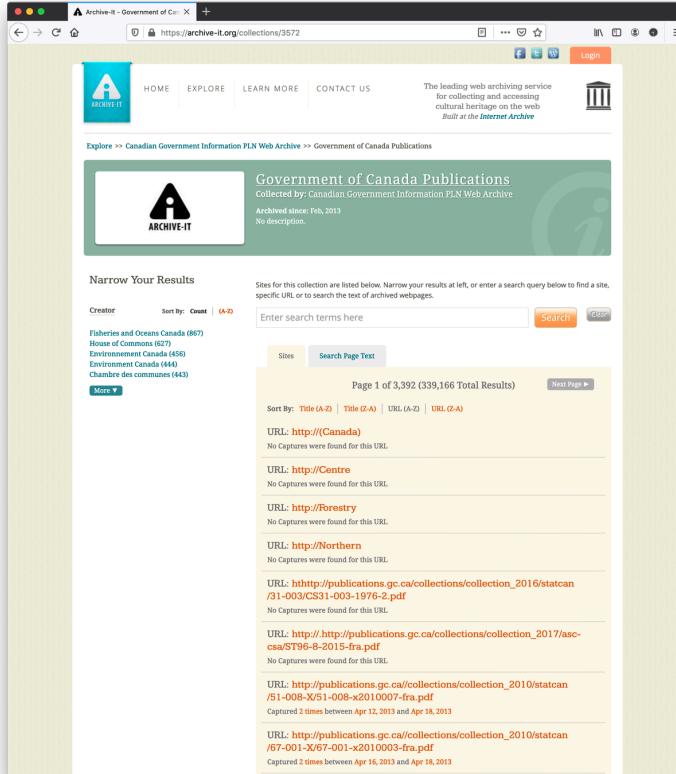


The screenshot shows the Archive-It Egypt Revolution collection interface. At the top, there's a navigation bar with links for HOME, EXPLORE, LEARN MORE, and CONTACT US. Below the navigation is a search bar with the placeholder 'Search the collection' and a 'Search' button. The main content area displays a collection of web pages from the Egypt Revolution and Politics category. Each item in the list includes a thumbnail image, the title, a brief description, and a 'View Details' link. The items are arranged in a grid format with three columns. The titles of the visible items include 'Egypt - Performance - Elections 2011', 'Egypt - Politics and government (123)', 'Egypt - Politics and government (111)', 'Politics and government (10)', 'Egypt - Politics and government (1)', 'Creator', 'Publisher', and 'Source'. The dates of the captured pages range from 2010 to 2015.

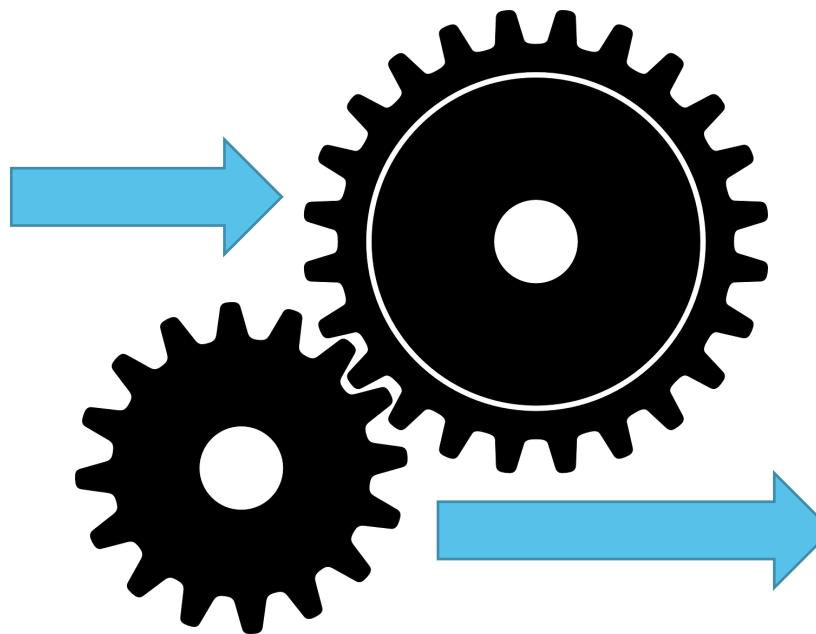


@shawnmjones @StormyArchives

# Ideally, a user would be able to glance at a visualization and gain understanding of the collection



Web Archive Collection



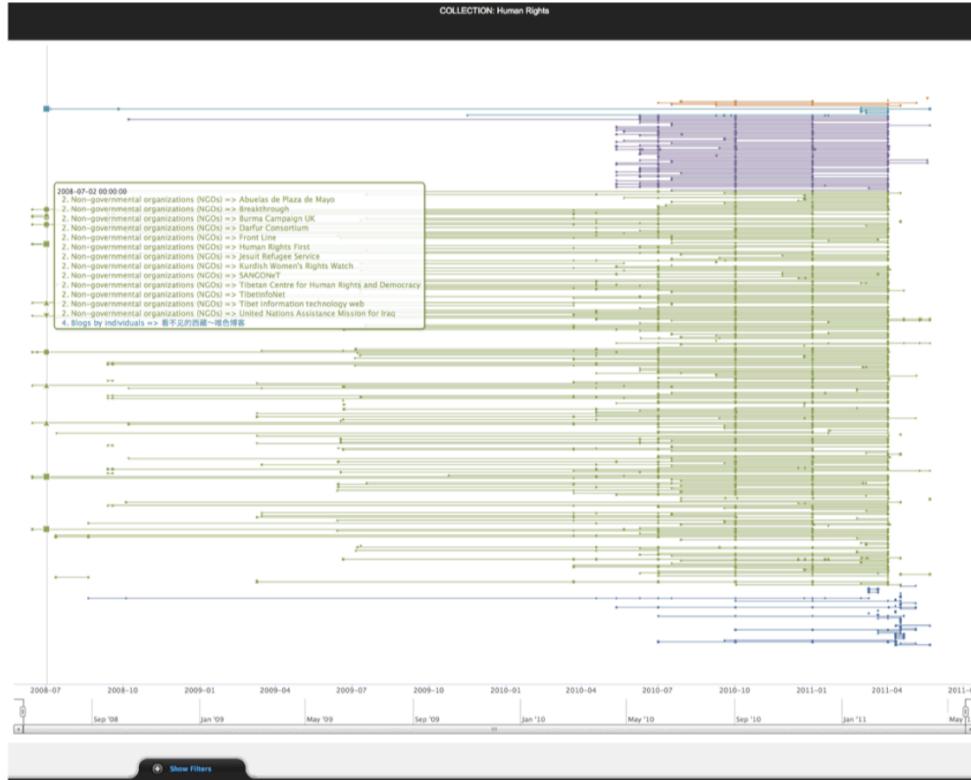
Automated  
Solution



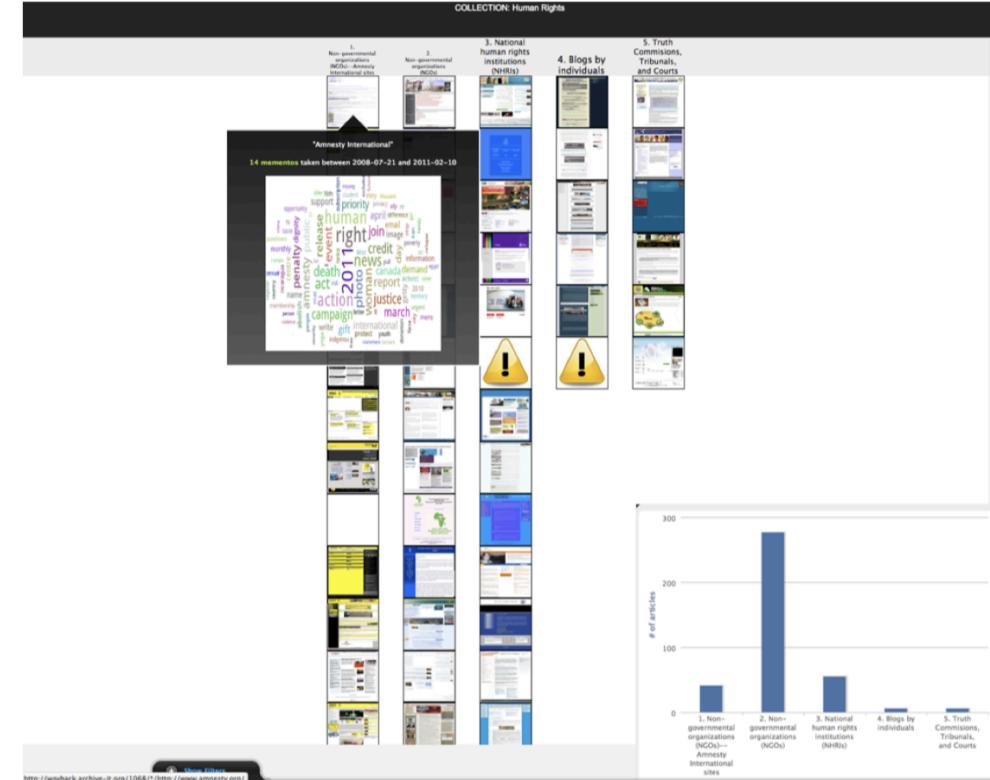
Visualization that  
conveys understanding  
at a glance



But techniques that apply to every page are hard to scale up and difficult to understand...



Timelines provide an overview, but do not provide understanding at a glance.



Thumbnails work well for small collections, but do not scale.



@shawnmjones @StormyArchives

K. Padia, Y. AlNoamany, and M. C. Weigle, (2012) "Visualizing digital collections at Archive-It." In JCDL '12, <https://doi.org/10.1145/2232817.2232821>

# What if we could apply a visualization technique that users already know how to process?

Users would require no training and could instead get information immediately.

Users could quickly decide which collection meets their needs!



# Social media storytelling



@shawnmjones @StormyArchives

The image contains three screenshots of social media platforms:

- Tumblr:** A post titled "Archive-It Stories" featuring a photo of two people in a flooded area. The caption reads: "Watertown Residents Hunker Down As Manhunt Unfolds".
- Archive-It Stories:** A post titled "Archive-It Stories" featuring a photo of a person in a flooded area. The caption reads: "Search for marathon bombing suspect leads down Watertown, surrounding communities".
- Boston Marathon Bombing 2013:** A post titled "Boston Marathon Bombing 2013" featuring a photo of a person in a flooded area. The caption reads: "Boston Marathon Bombing 2013".

# Web surrogates provide a visual summary of a web resource drawn from the content of the resource

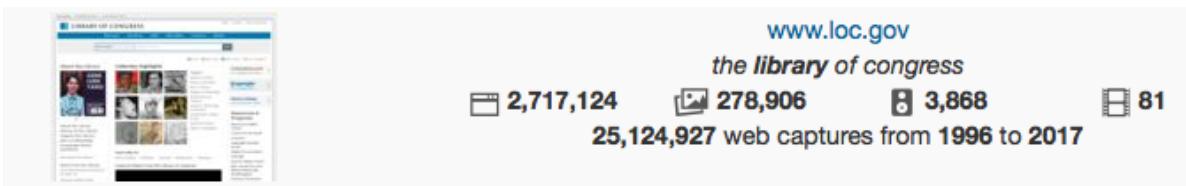
[Home](#) | [Michele C. Weigle](#)

[www.cs.odu.edu/~mweigle](http://www.cs.odu.edu/~mweigle) ▾

Michele C. Weigle, Michael L. Nelson, Deborah Kempe (Frick Art Reference Library and New York Art Resources Consortium), Pamela Graham (Columbia University Libraries), and Alex Thurman (Columbia University Libraries), "Visualizing Webpage Changes Over Time", NEH/IMLS Digital Humanities Advancement Grant HAA-256368-17, Oct 2017 – ...

[Research Methods](#) · [Publications](#) · [Schedule/Travel](#) · [Teaching](#) · [Working With Me](#)

Text snippet (example from Bing)



Text + Thumbnail (example from Internet Archive)



## Instances



Instances archived after 01-jul-2008 and before 31-jan-2010



Browser Thumbnail (example from UK Web Archive)



2017-08-11: Where Can We Post Stories  
Summarizing Web Archive Collections?

A social card generated by Facebook for my previous blog post. Rich links, snippet, social snippet, social media card, Twitter car...

WS-DL.BLOGSPOT.COM

Social Card (example from Facebook)

# Social cards are summaries of web resources

Long URL: <https://www.google.com/maps/dir/Old+Dominion+University,+Norfolk,+VA/Los+Alamos+National+Laboratory,+New+Mexico/@35.3644614,-109.356967,4z/data=!3m1!4b1!4m13!4m12!1m5!1m1!1s0x89ba99ad24ba3945:0xcd2bdc432c4e4bac!2m2!1d-76.3067676!2d36.8855515!1m5!1m1!1s0x87181246af22e765:0x7f5a90170c5df1b4!2m2!1d-106.287162!2d35.8440582>

VS.

The same URL represented as a social card:



- URLs can be difficult for people to comprehend
- Social cards provide a title, small text snippet, and striking image from the web page behind a URL



# And we use these cards all the time!



WS-DL.BLOGSPOT.COM

2017-08-11: Where Can We Post Stories Summarizing Web Archive Collections?



2017-08-11: Where Can We Post Stories Summarizing Web ...

The Web Science and Digital Libraries Research Group at Old Dominion University.

ws-dl.blogspot.com



ws-dl.blogspot.com

2017-08-11: Where Can We Post Stories Summarizing Web Archive Collections?

2017-08-11: Where Can We Post Stories Summarizing Web Archive Collections? >

The Web Science and Digital Libraries Research Group at Old Dominion University.

## Facebook

2017-08-11: Where Can We Post Stories Summarizing Web ...

<https://ws-dl.blogspot.com/2017/08/2017-08-11-where-can-we-post-stories.html> ▾

Aug 11, 2017 · In Yasmin AlNoamany's Dark and Stormy Archives (DSA), she uses social cards in Storify stories to summarize mementos from Archive-It collections. These stories take the form of 28 high ...

Author: Shawn M. Jones

## Twitter

ws-dl.blogspot.com > 2017/08 > 2017-08-11-where-ca... ▾

Where Can We Post Stories Summarizing Web Archive ...

Aug 11, 2017 – In effect, social cards are visualizations of web objects, piercing the veil ... Earlier attempts at visualizing Archive-It collections by Padia and ...

## Bing

## Google



@shawnmjones @StormyArchives

# Social media storytelling uses groups of social cards to provide a “summary of summaries”

Boston Marathon Bombing 2013

by art\_stories 8 months ago 60 Views ~

**Watertown Residents Hunker Down As Manhunt Unfolds : NPR**  
NPR's Matt O'Brien is a former NPR contributor who lives and works near the scene of the overnight showdown in Watertown, Mass. While the manhunt for a suspect in the Boston Marathon bombings continued Thursday night into Friday morning, residents of Watertown, Mass., and surrounding communities were hiding in bedrooms, looking out from roofs and peering from behind locked doors.

**Search for marathon bombing suspect locks down Watertown, surrounding communities**  
With nearly a million residents of the Boston area hunkering down behind locked doors, heavily armed police are making their way through a 20-block area of Watertown, searching for one of the suspects in the deadly Boston Marathon terror bomb attacks, the head of the State Police said this afternoon.

**19. Apr 2013 - Boston Marathon Explosions: Thursday's Developments : The Two-Way : NPR**  
"The grace this tragedy exposed is the best of who we are," Patrick added just before introducing President Obama. The cardinal is to be followed at the interfaith service by a performance from cellist Yo-Yo Ma. Thursday's post: throughout the day, we'll be updating with the latest news about the tw...

**President Obama: "The American People Refuse to be Terrorized" | The White House**  
Following a briefing from FBI Director Mueller, Attorney General Holder, Secretary Napolitano, and homeland security advisor Lisa Monaco, President Obama went to the Brady Press Briefing Room to update Americans on developments in Boston, following two explosions at the finish line of the Boston Marathon on Monday afternoon.

**18. Apr 2013 - Obamas join Boston at memorial for bombing victims**  
Officers are trawling through hotel bookings and short-term accommodation rentals throughout the Boston area for any clues. Homeland security secretary Janet Napolitano said the FBI wanted to interview two men seen in at least one video. Pieces of what is believed to be a circuit board have also bee...

**Reddit and 4chan Investigate Boston Marathon Attack**  
As government investigators in Boston and around the country have been working to follow the winding stream of forensic data and digital imagery toward the terrorist behind the Boston Marathon bombing, a parallel investigation has been happening, completely in the open, on the

Each social card summarizes a web resource.

Each story groups the social cards, summarizing the topic.

Social cards contain the same information in the same place on each card, allowing for easy comparison.

We want to use this technique to summarize web archive collections **because users are already familiar with this visualization paradigm.**

6 resources are shown in this Storify story



@shawnmjones @StormyArchives

W A Hunger Crisis - Wakelet

wakelet

**Created by The Rohingya Crisis**

**A Hunger Crisis**  
16 items

As thousands are fleeing to neighbouring Bangladesh, they are faced with overfilled refugee camps where food remains scarce. Many families face extreme hunger and malnutrition, especially the weak and children.

**Myanmar accused of using starvation against Rohingya**  
Aljazeera  
UN high commissioner for human rights says Myanmar is now using starvation as a weapon against the Rohingya after months of violent crackdowns.

**UN says 14,000 Rohingya children may die from malnutrition**  
Aljazeera  
As the Rohingya exodus from Myanmar nears 600,000 people, the United Nations has issued a new report on children caught in the crisis.

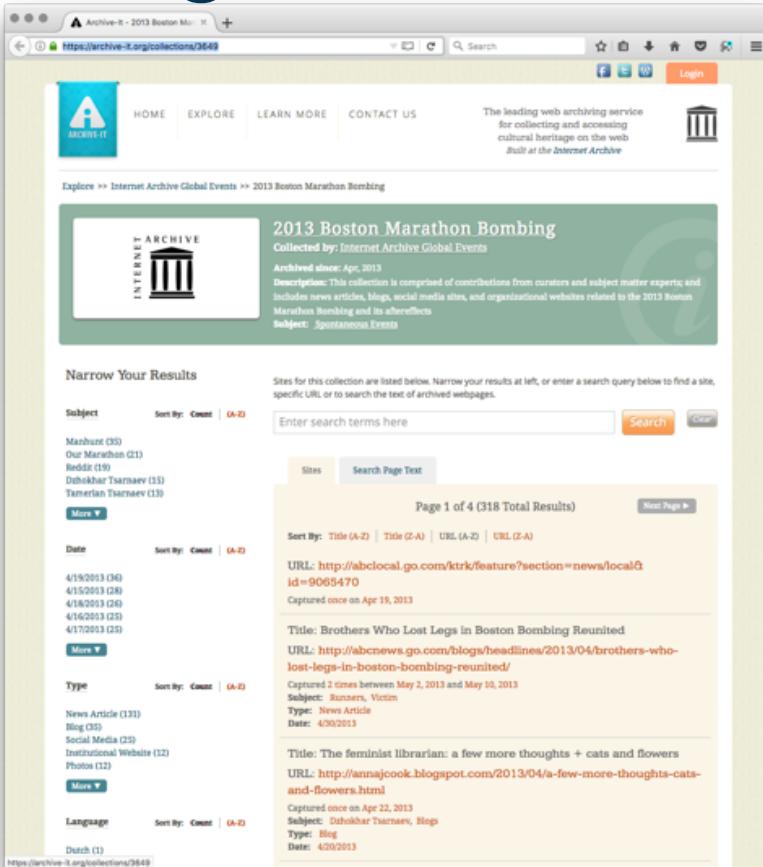
2 resources are shown in this Wakelet story

**Social media storytelling uses an interface that users already understand how to process!**

Users would require no training and could instead get information immediately.

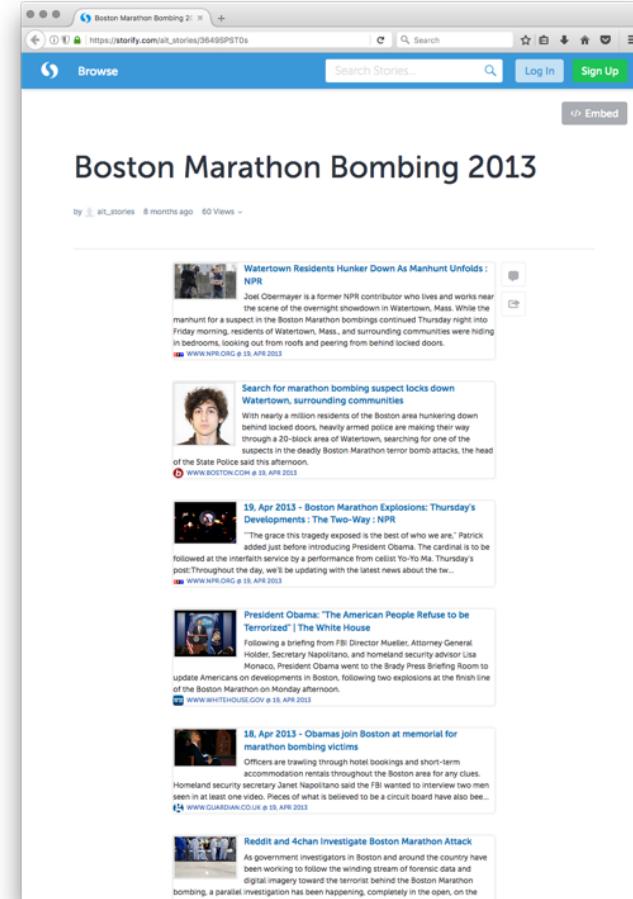
Users could quickly decide which collection meets their needs!

# Our solution: applying social media storytelling to web archives



Archive-It interface showing search results for the 2013 Boston Marathon Bombing. The results are filtered by 'Subject' and 'Date'. One result is highlighted with a blue box, showing a URL and capture date.

From this:  
318 seeds with  
2421 mementos

Boston Marathon Bombing 2013

Watertown Residents Hunker Down As Manhunt Unfolds : NPR  
Joet Obermayer is a former NPR contributor who lives and works near the scene of the overnight showdown in Watertown, Mass. While the manhunt for a suspect in the Boston Marathon bombings continued Thursday night into Friday morning, residents of Watertown, Mass., and surrounding communities were hiding in bedrooms, looking out from roofs and peering from behind locked doors.

Search for marathon bombing suspect locks down Watertown, surrounding communities  
With nearly a million residents of the Boston area hunkering down behind locked doors, heavily armed police are making their way through the surrounding area of Watertown, searching for one of the suspects in the deadly Boston Marathon terror bomb attacks, the head of the State Police said this afternoon.

19. Apr 2013 - Boston Marathon Explosions: Thursday's Developments : The Two-Way : NPR  
"The grace this tragedy exposed is the best of who we are," Patrick added just before introducing President Obama. The cardinal is to be followed at the interfaith service by a performance from cellist Yo-Yo Ma. Thursday's post: Throughout the day, we'll be updating with the latest news about the tw...

President Obama: "The American People Refuse to be Terrorized" | The White House  
Following a briefing from FBI Director Mueller, Attorney General Holder, Secretary Napolitano, and homeland security advisor Lisa Monaco, President Obama and First Lady Michelle Obama went to the Brady Press Briefing Room to update Americans on developments in Boston, following two explosions at the finish line of the Boston Marathon on Monday afternoon.

18. Apr 2013 - Obamas join Boston at memorial for marathon bombing victims  
Officers are traveling through hotel bookings and short-term accommodation rentals throughout the Boston area for any clues. Homeland security secretary Janet Napolitano said the FBI wanted to interview two men seen in at least one video. Pieces of what is believed to be a circuit board have also bee...

Reddit and 4chan Investigate Boston Marathon Attack  
As government investigators in Boston and around the country have been working to follow the winding stream of forensic data and digital imagery toward the terrorist behind the Boston Marathon bombing, a parallel investigation has been happening, completely in the open, on the

To something like this:  
a **social media story**  
of ~28 cards



@shawnmjones @StormyArchives

# Storytelling With Web Archives Has Multiple Use Cases...

## 1. Promotion of the collection

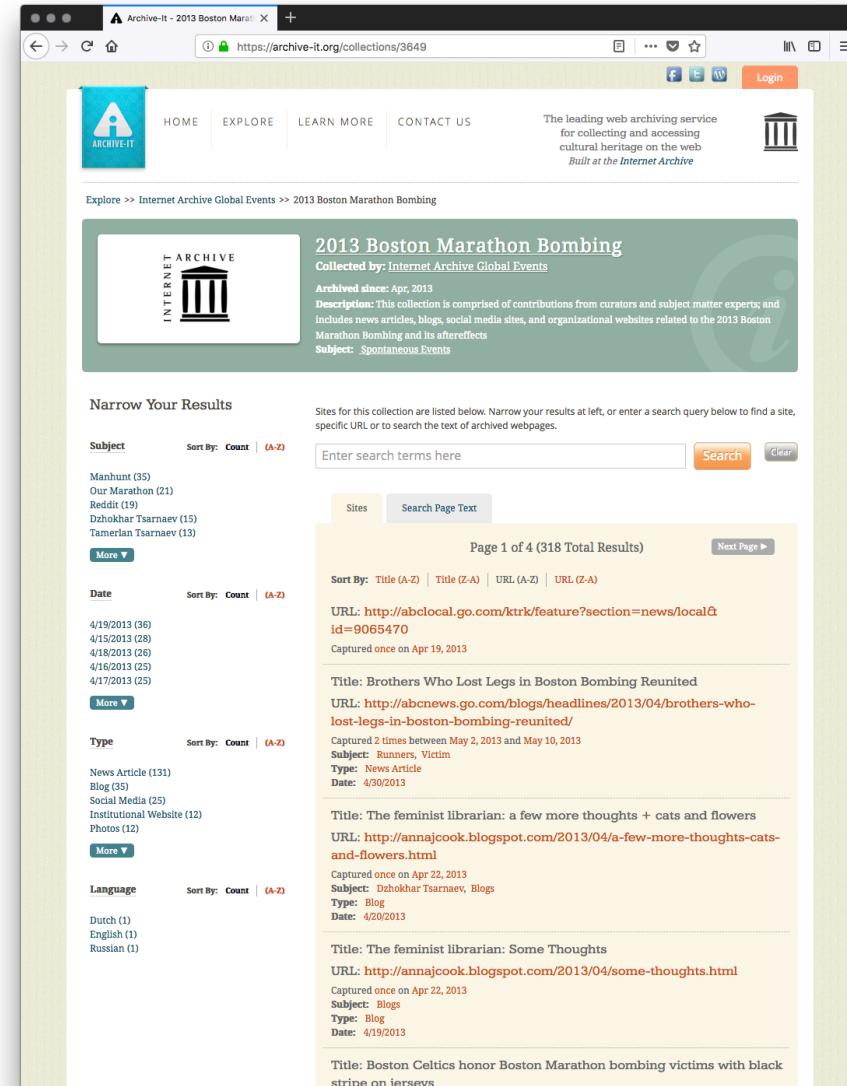
- Storytelling allows a curator to promote a collection, making others aware of it

## 2. Exploring aspects of the collection

- Users can explore a collection and expose specific sides of a news story or focus on specific people or places

## 3. Summarization

- Web archive collections are too large for manual review – we need a summary to understand what they contain



The screenshot shows the Archive-It collection page for the 2013 Boston Marathon Bombing. The header includes the Archive-It logo, navigation links (HOME, EXPLORE, LEARN MORE, CONTACT US), and a sub-navigation link for 'Internet Archive Global Events'. Below the header, a banner for the '2013 Boston Marathon Bombing' collection is displayed, noting it was collected by 'Internet Archive Global Events' and archived since April 2013. The collection description includes contributions from curators and subject matter experts, news articles, blogs, social media sites, and organizational websites related to the event. The main content area is titled 'Narrow Your Results' and lists filters for 'Subject', 'Date', 'Type', and 'Language'. The 'Subject' filter shows categories like Manhunt (35), Our Marathon (21), Reddit (19), Dzhokhar Tsarnaev (15), and Tamerlan Tsarnaev (13). The 'Date' filter shows dates from April 19, 2013, to April 17, 2013. The 'Type' filter shows News Article (131), Blog (35), Social Media (25), Institutional Website (12), and Photos (12). The 'Language' filter shows Dutch (1), English (1), and Russian (1). The results list shows four items, each with a title, URL, capture date, subject, type, and date. The first item is 'Brothers Who Lost Legs in Boston Bombing Reunited' from abclocal.go.com, captured on April 19, 2013. The second item is 'The feminist librarian: a few more thoughts + cats and flowers' from annajcook.blogspot.com, captured on April 22, 2013. The third item is 'Boston Celtics honor Boston Marathon bombing victims with black stripe on jerseys' from annajcook.blogspot.com, captured on April 19, 2013.



# Some services equate “stories” with ephemera, which is not what we want



Facebook stories

Image ref:

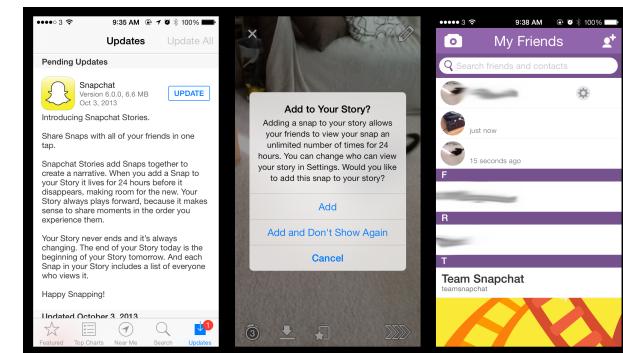
<https://techcrunch.com/2018/04/05/facebook-stories-default/>



Instagram stories

Image ref:

<https://buffer.com/library/instagram-stories>



Snapchat stories

Image ref:

<https://techcrunch.com/2013/10/03/snapchat-gets-its-own-timeline-with-snapchat-stories-24-hour-photo-video-tales/>

These platforms delete the user's stories 24 hours after they are posted.

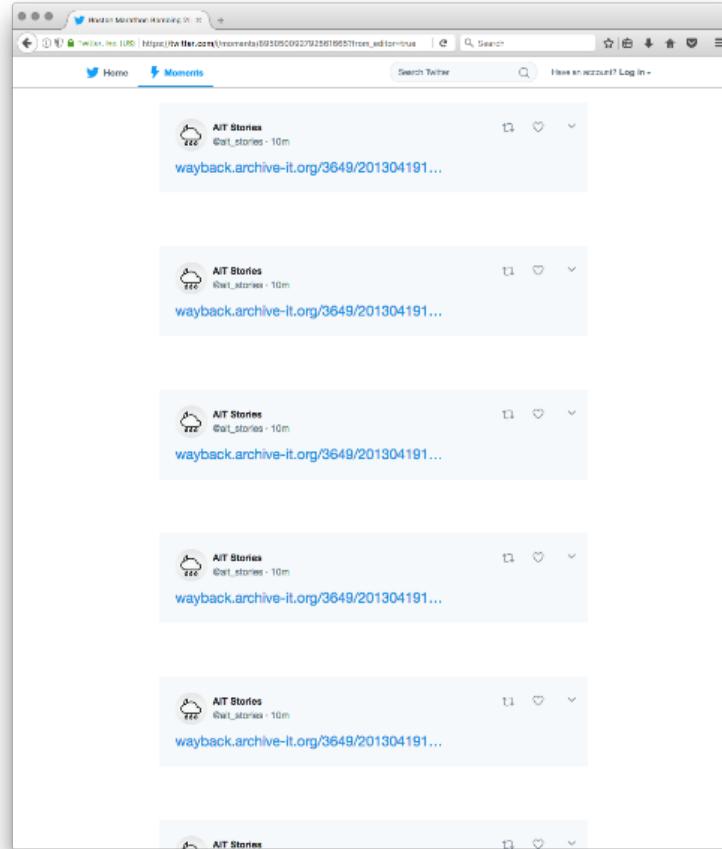
This form of social media storytelling is the opposite of what we are looking for.  
We want the stories to be artifacts themselves.



@shawnmjones @StormyArchives

Jones, S.M., Klein, M., Weigle, M.C., & Nelson, M.L. (2020). MementoEmbed and Raintale for Web Archive Storytelling. In Web Archiving and Digital Libraries Workshop (WADL 2020). arXiv, abs/2008.00137. <https://arxiv.org/abs/2008.00137>

# Existing social platforms do not work well for archived resources (mementos)



 **Archive-It Stories** <http://wayback.archive-it.org/.../www.wcvb.com/news/local>  
August 3 at 5:26pm

 **Archive-It Stories** <http://wayback.archive-it.org/.../8uoVeyShelDu.../story.html>  
  
Patriot's Day heroes: Marathon spectators become first- responders - The Boston...  
BOSTONGLOBE.COM

August 3 at 5:27pm



# Existing card services create a confusing experience for pages from web archives (mementos)

CN ARCHIVE-IT



Egypt - This Just In - CNN.com Blogs

To alleviate U.S. pressure on the Egyptian army, Saudi Arabia would consider matching the estimated \$1.5 billion the United States provides Egypt in military aid each year if Washington decides to cut its aid to the troubled North African nation, an Arab diplomat says.

[Read the article on wayback.archive-it.org >](#)

powered by embedly

embed.ly social card

This is confusing.

Who published these resources?  
Archive-It?  
CNN?

It is a matter of trust, because different sources are attributed.



@shawnmjones @StormyArchives

Jones, S.M., Klein, M., Weigle, M.C., & Nelson, M.L. (2020). MementoEmbed and Raintale for Web Archive Storytelling. In Web Archiving and Digital Libraries Workshop (WADL 2020). arXiv, abs/2008.00137. <https://arxiv.org/abs/2008.00137>

# Current social media storytelling solutions don't work with archival content

Our stories should not be temporary.

Our summaries should not present users with  
confusing attribution.

Web archive collections  
are vast and hard to  
understand

Understanding them is  
difficult with existing  
visualizations

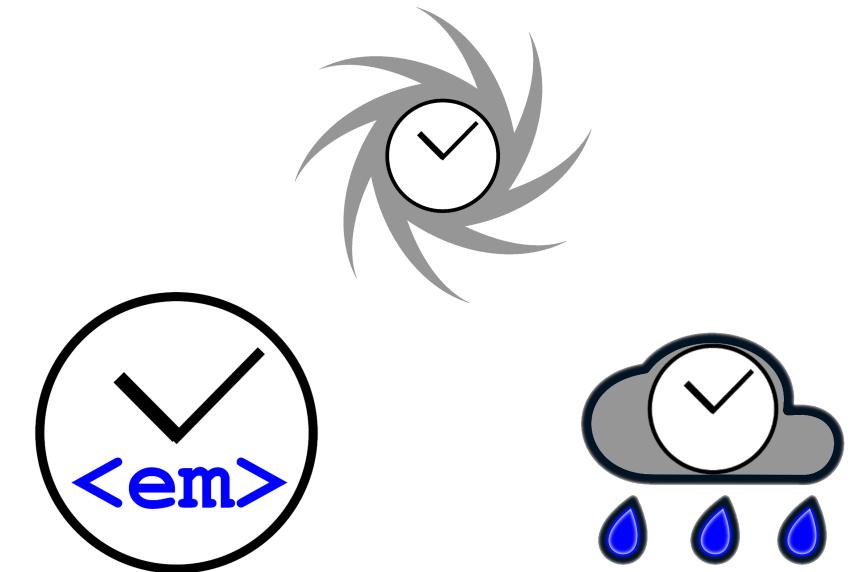
Existing social media  
storytelling does a  
poor job with archived  
pages (mementos)



# Solving the problem with the Dark and Stormy Archives Project



@shawnmjones @StormyArchives



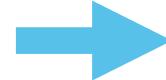
# Telling a story with web archives has 3 basic steps

1. Select the pages  
for the story

2. Gather information  
to summarize  
each memento

3. Summarize  
all mementos together  
and publish the story

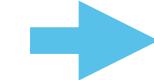
?



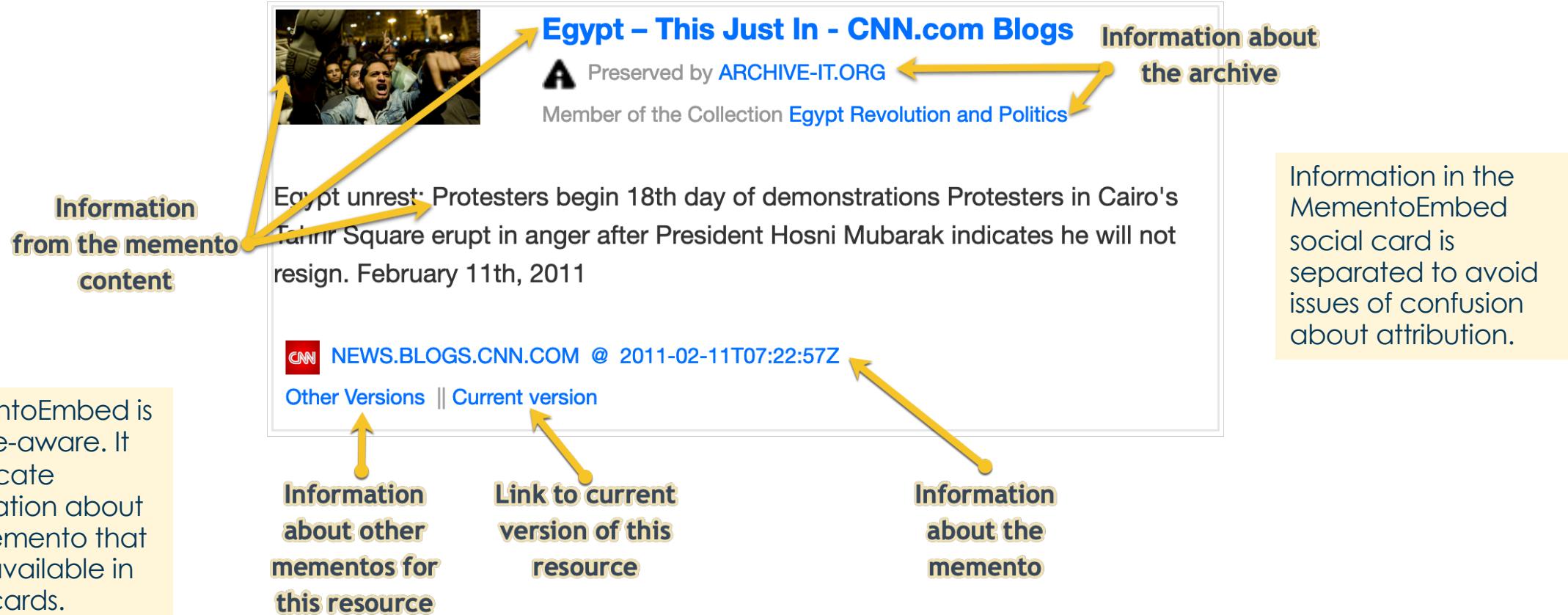
?



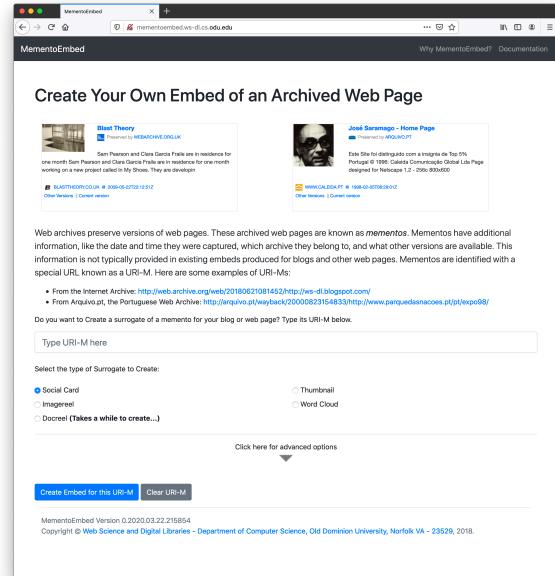
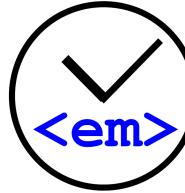
?



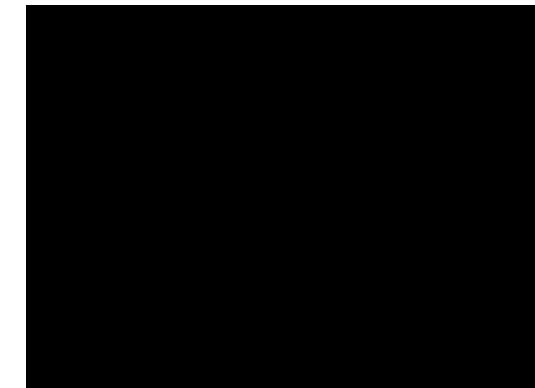
# Neither social media services nor card services were reliable for web archive storytelling, so we created MementoEmbed...



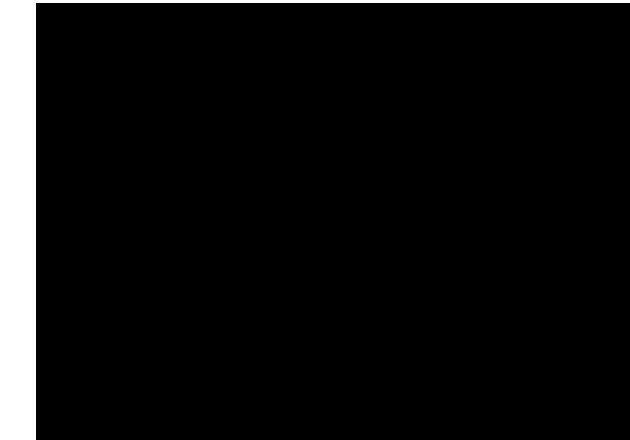
We can summarize *individual* pages with MementoEmbed in different ways



All of these summaries represent the same page about the Egyptian Revolution.



Imagereels



Docreels



@shawnmjones @StormyArchives

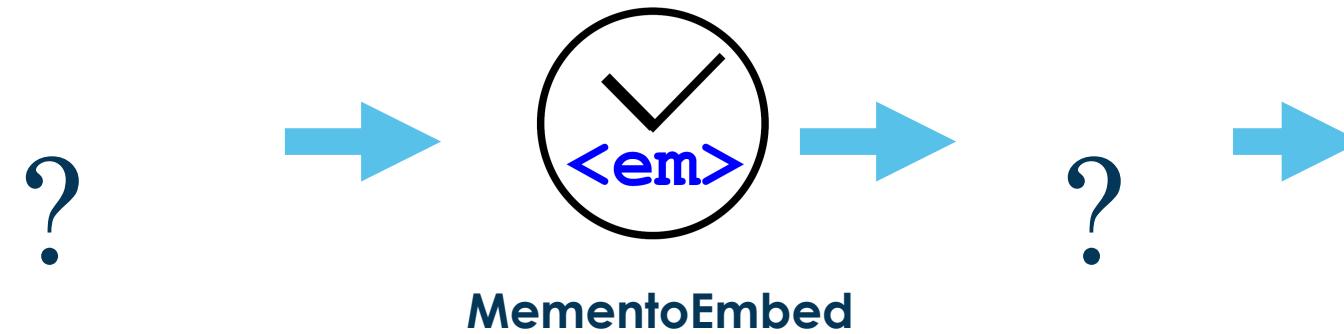


# Now we can produce cards, but how do we create stories?

1. Select the pages  
for the story

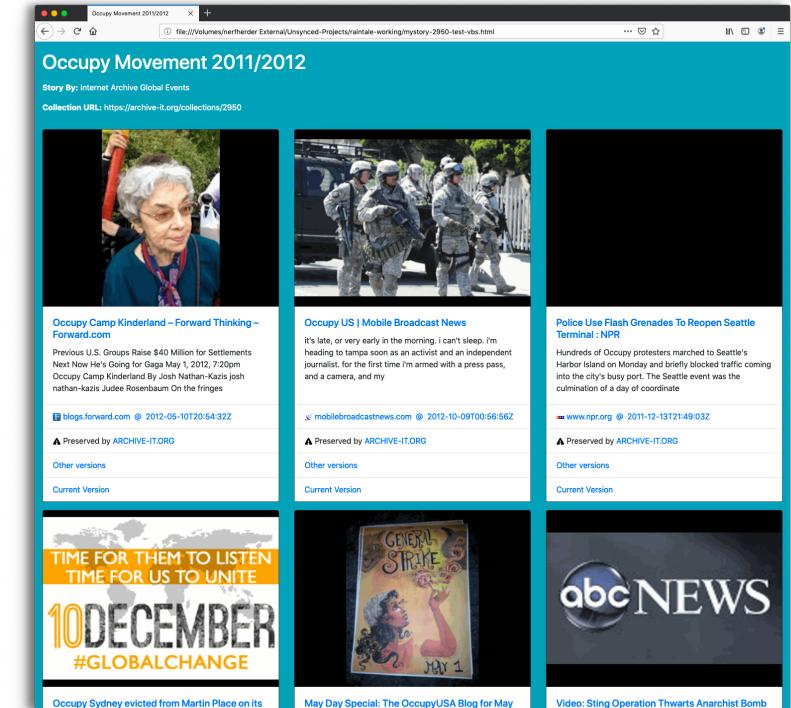
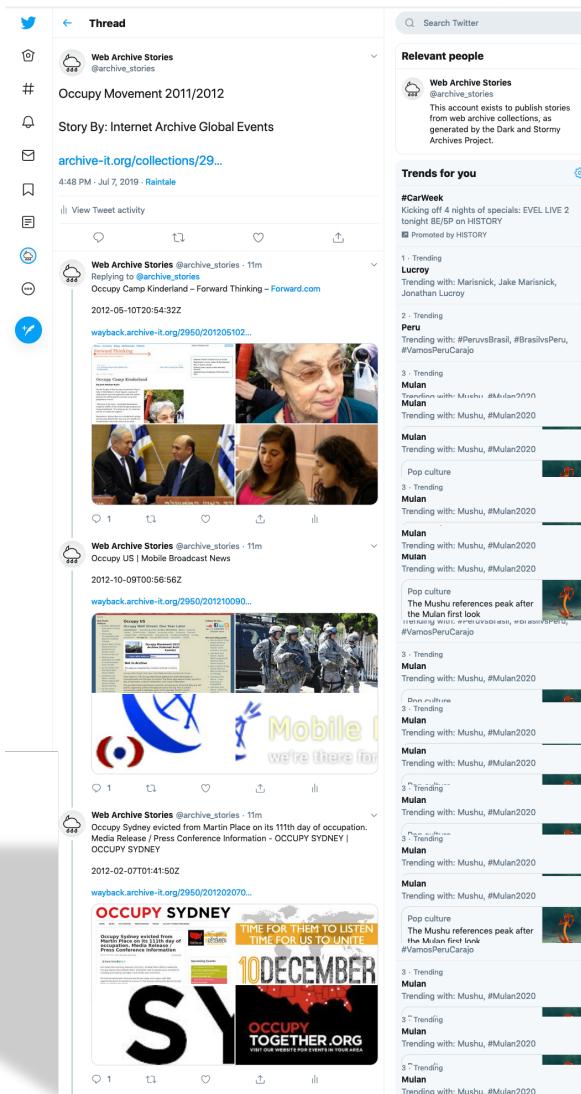
2. Gather information  
to summarize  
each memento

3. Summarize  
all mementos together  
and publish the story



# MementoEmbed just produces information for individual pages, so we created Raintale to tell stories...

- Raintale generates social media stories for mementos
- Raintale publishes to file formats, like HTML, or social media services, like Twitter
- Raintale also supports templates for custom output



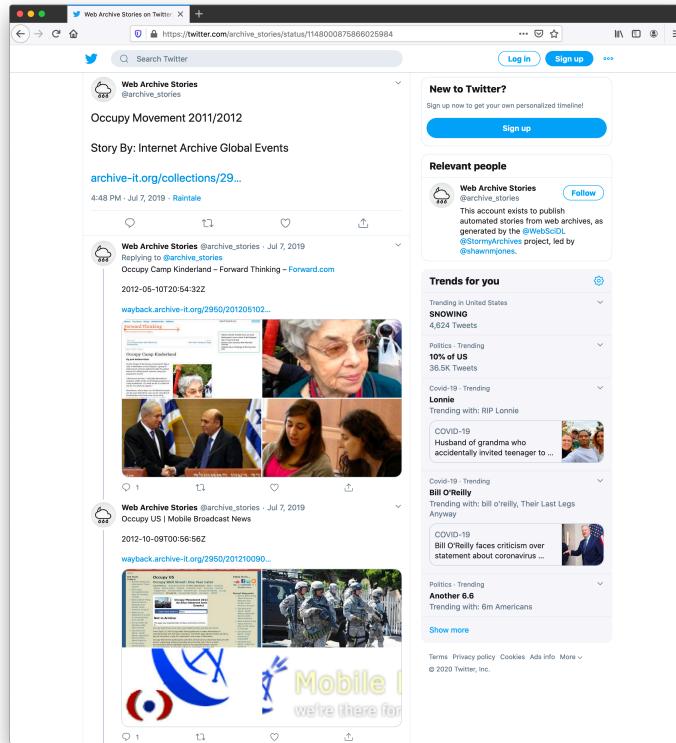
<https://oduwsdl.github.io/raintale/>



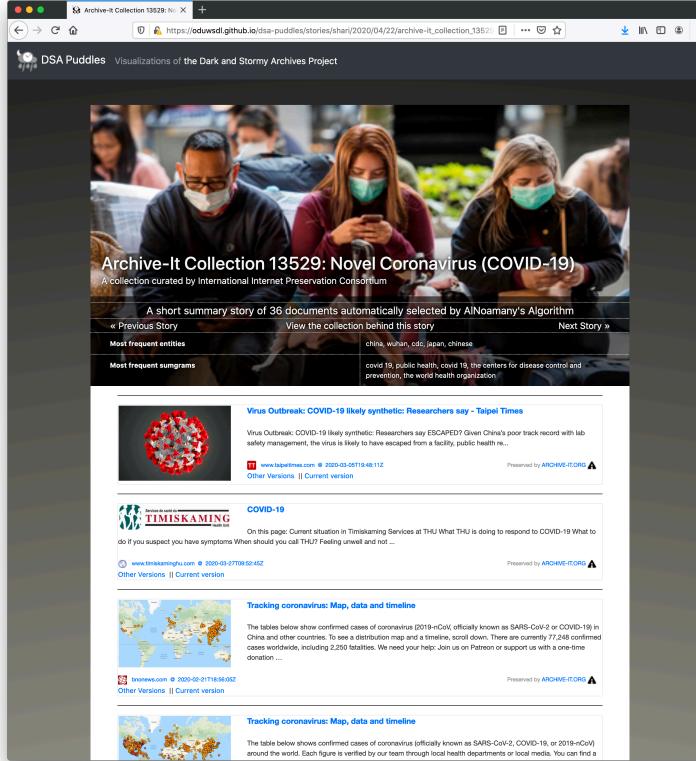
@shawnmjones @StormyArchives

Jones, S.M., Klein, M., Weigle, M.C., & Nelson, M.L. (2020). MementoEmbed and Raintale for Web Archive Storytelling. In Web Archiving and Digital Libraries Workshop (WADL 2020). arXiv, abs/2008.00137. <https://arxiv.org/abs/2008.00137>

# We can publish *many* Mementos with Raintale in a variety of formats



Twitter Threads  
with screenshots,  
images, titles, dates, links



GitHub Pages  
with cards and much more



Occupy Movement 2011/2012

Generated by Internet Archive Global Events

Videos  
with top ranked images and sentences  
to be uploaded to YouTube, Twitter, etc.  
(experimental)



@shawnmjones @StormyArchives

Jones, S.M., Klein, M., Weigle, M.C., & Nelson, M.L. (2020). MementoEmbed and Raintale for Web Archive Storytelling. In Web Archiving and Digital Libraries Workshop (WADL 2020). arXiv, abs/2008.00137. <https://arxiv.org/abs/2008.00137>

# Raintale uses templates to let you customize the look and destination media of your story



Web Archive Stories

Story By: Collection Archive

Collection URL: <https://archive-it.org/collections/2950>

Occupied Camp Kinderland - Forward Thinking - Forward.com

Occupied Camp Kinderland - Forward Thinking - Forward.com

Occupied US | Mobile Broadcast News

Occupied US | Mobile Broadcast News

Police Use Flash Grenades To Reopen Seattle Terminal - NPR

Police Use Flash Grenades To Reopen Seattle Terminal - NPR

Occupy Sydney | Occupy Sydney evicted from Martin Place on its 111th day of occupation. Media Release / Press Conference Information - OCCUPY SYDNEY | OCCUPY SYDNEY

Occupy Sydney | Occupy Sydney evicted from Martin Place on its 111th day of occupation. Media Release / Press Conference Information - OCCUPY SYDNEY | OCCUPY SYDNEY

May Day Special: The OccupyUSA Blog for May 1, With Frequent Updates! | The Nation

May Day Special: The OccupyUSA Blog for May 1, With Frequent Updates! | The Nation

Occupied Movement 2011/2012

Story By: Internet Archive Global Events

Collection URL: <https://archive-it.org/collections/2950>

Occupied Camp Kinderland - Forward Thinking - Forward.com

Occupied Camp Kinderland - Forward Thinking - Forward.com

Occupied US | Mobile Broadcast News

Occupied US | Mobile Broadcast News

Police Use Flash Grenades To Reopen Seattle Terminal - NPR

Police Use Flash Grenades To Reopen Seattle Terminal - NPR

Occupy Sydney | Occupy Sydney evicted from Martin Place on its 111th day of occupation. Media Release / Press Conference Information - OCCUPY SYDNEY | OCCUPY SYDNEY

Occupy Sydney | Occupy Sydney evicted from Martin Place on its 111th day of occupation. Media Release / Press Conference Information - OCCUPY SYDNEY | OCCUPY SYDNEY

May Day Special: The OccupyUSA Blog for May 1, With Frequent Updates! | The Nation

May Day Special: The OccupyUSA Blog for May 1, With Frequent Updates! | The Nation

Occupy Movement Story

Occupy Movement 2011/2012 [edit]

Story By: Internet Archive Global Events

Collection URL: <https://archive-it.org/collections/2950>

Occupied Camp Kinderland - Forward Thinking - Forward.com

Occupied Camp Kinderland - Forward Thinking - Forward.com

Occupied US | Mobile Broadcast News

Occupied US | Mobile Broadcast News

Police Use Flash Grenades To Reopen Seattle Terminal : NPR

Police Use Flash Grenades To Reopen Seattle Terminal : NPR

Occupy Sydney | Occupy Sydney evicted from Martin Place on its 111th day of occupation. Media Release / Press Conference Information - OCCUPY SYDNEY | OCCUPY SYDNEY

Occupy Sydney | Occupy Sydney evicted from Martin Place on its 111th day of occupation. Media Release / Press Conference Information - OCCUPY SYDNEY | OCCUPY SYDNEY

May Day Special: The OccupyUSA Blog for May 1, With Frequent Updates! | The Nation

May Day Special: The OccupyUSA Blog for May 1, With Frequent Updates! | The Nation

My Archive About Contact

Occupied Movement 2011/2012

Occupied Camp Kinderland - Forward Thinking - Forward.com

Occupied US | Mobile Broadcast News

Police Use Flash Grenades To Reopen Seattle Terminal : NPR

Occupy Sydney | Occupy Sydney evicted from Martin Place on its 111th day of occupation. Media Release / Press Conference Information - OCCUPY SYDNEY | OCCUPY SYDNEY

May Day Special: The OccupyUSA Blog for May 1, With Frequent Updates! | The Nation

MementoEmbed  
cards in Blogger

Customized Twitter  
Threads

MediaWiki  
Pages

Brand your own stories  
for your institution



@shawnmjones @StormyArchives

Jones, S.M., Klein, M., Weigle, M.C., & Nelson, M.L. (2020). MementoEmbed and Raintale for Web Archive Storytelling. In Web Archiving and Digital Libraries Workshop (WADL 2020). arXiv, abs/2008.00137. <https://arxiv.org/abs/2008.00137>

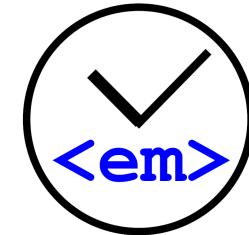
# Now we can create stories, but how do we select mementos?

1. Select the pages  
for the story

2. Gather information  
to summarize  
each memento

3. Summarize  
all mementos together  
and publish the story

?



MementoEmbed

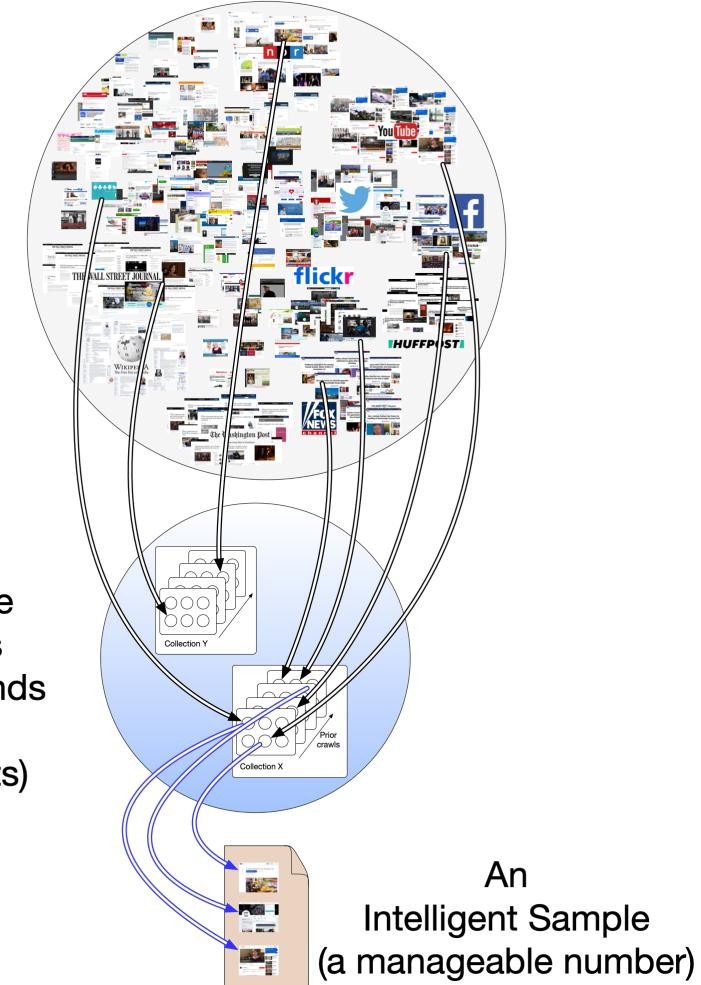


Raintale

# Hypercane provides intelligent sampling of web archive collections

- Works with Memento-compliant archives
- Discovers mementos
- Can create new mementos
- Reports on:
  - collection metadata
  - named entities
  - terms
  - image analysis
- Provides summarization steps:
  - Clustering
  - Filtering
  - Scoring
  - Ordering

The Web in the “perpetual now”  
(near  $\infty$  documents)



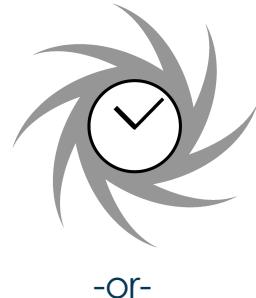
@shawnmjones @StormyArchives

S. M. Jones. “Hypercane Part 1: Intelligent Sampling of Web Archive Collections” <https://ws-dl.blogspot.com/2020/06/2020-06-03-hypercane-part-1-intelligent.html>, 2020.

# The DSA Toolkit provides a solution for each stage of the storytelling lifecycle

Automatically  
with  
**Hypercane**

1. Select the pages  
for the story

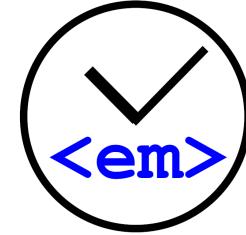


-or-

By People



2. Gather information  
to summarize  
each memento

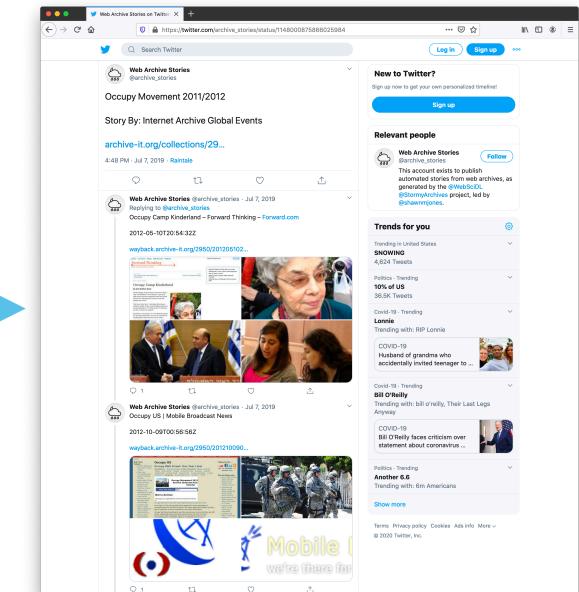


**MementoEmbed**

3. Summarize  
all mementos together  
and publish the story



**Raintale**



@shawnmjones @StormyArchives



# Examples of the DSA Toolkit at work

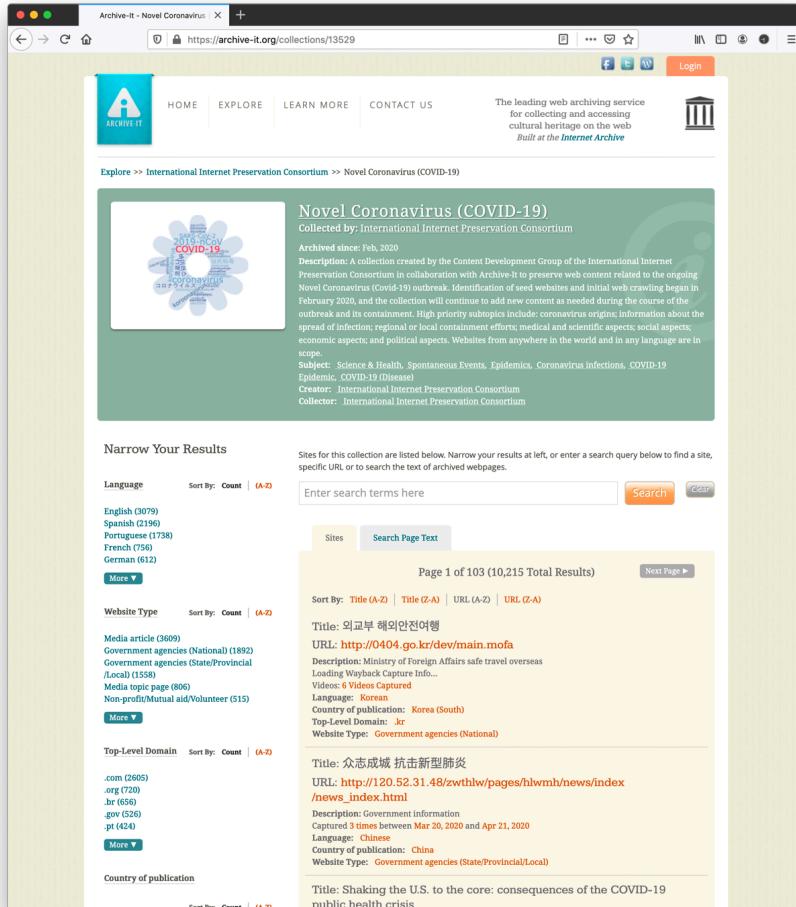
This screenshot shows a web-based visualization for the "Archive-It Collection 13529: Novel Coronavirus (COVID-19)". The main content area features a photograph of several people wearing face masks. Below the image, a summary text reads: "A short summary story of 36 documents automatically selected by AIComany's Algorithm". It includes links for "Previous Story", "Next Story", and "View the collection". The page is titled "DSA Puddles" and includes a "StoryGraph Biggest Story 2020-07-18" section.

This screenshot shows a web-based visualization for the "StoryGraph Biggest Story 2020-07-18" related to the civil rights movement. It features a photograph of John Lewis, Martin Luther King Jr., and other civil rights leaders. The main content area includes a summary of the story and links to "Previous Story", "Next Story", and "View the Storygraph". The page is titled "DSA Puddles" and includes a "StoryGraph Biggest Story 2020-07-18" section.

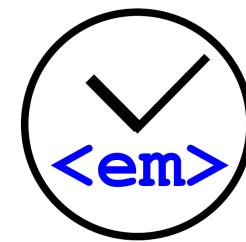
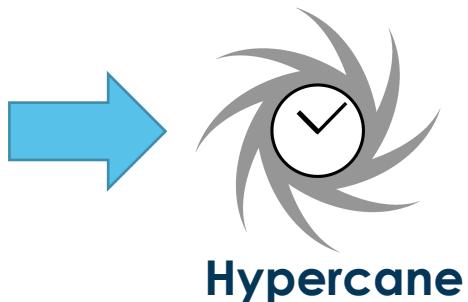


@shawnmjones @StormyArchives

# We automatically summarize public Archive-It collections



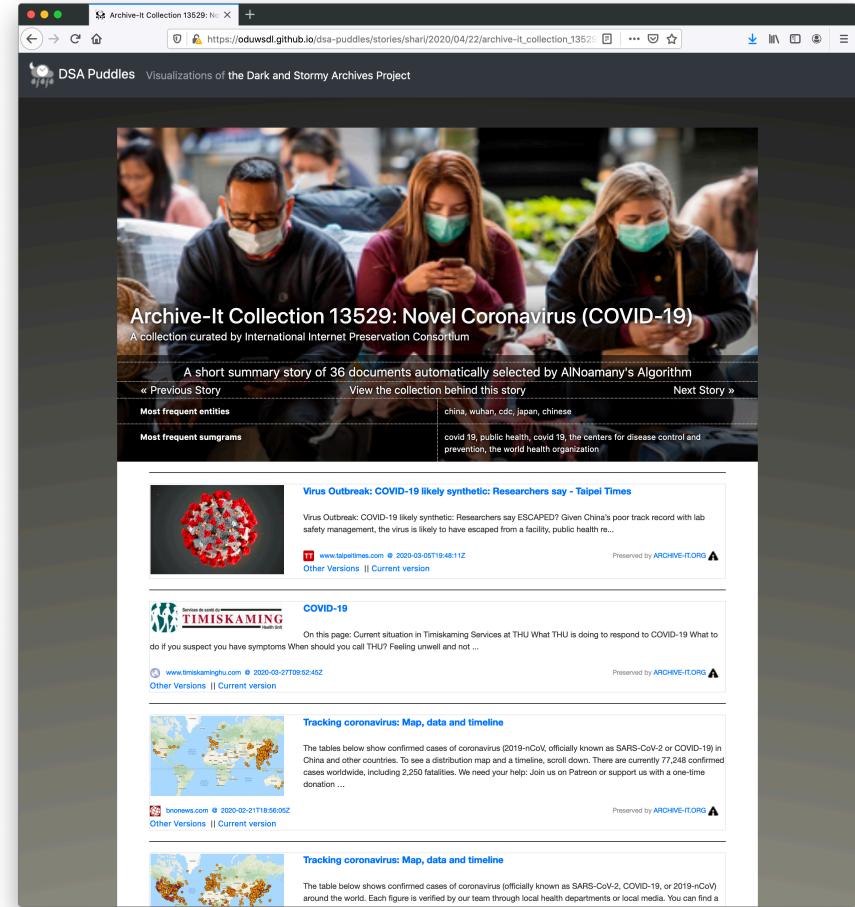
From this: 23,376 mementos



MementoEmbed



Raintale



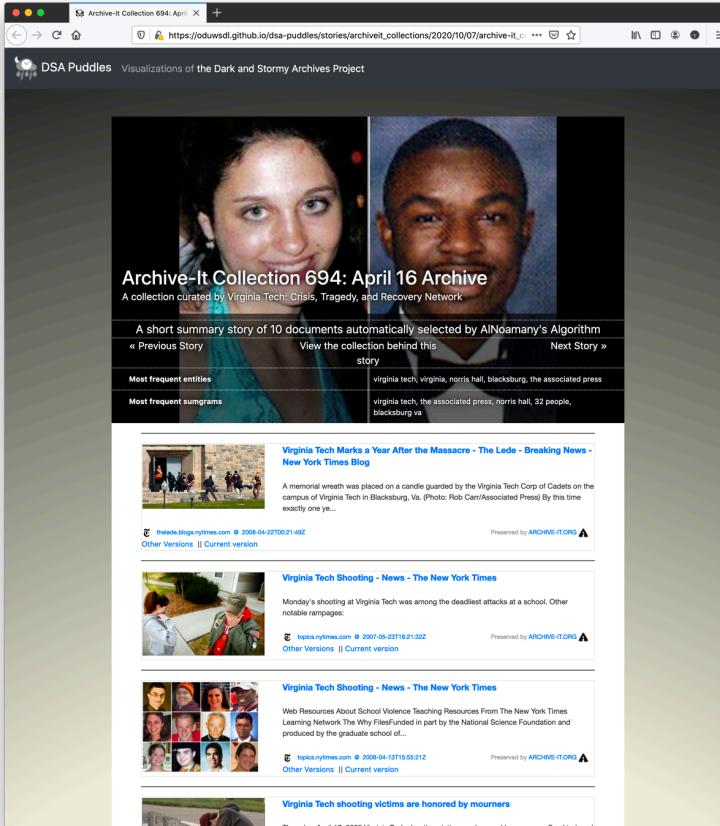
To this: a sample of 36 mementos visualized as social cards, phrases, and images



@shawnmjones @StormyArchives

Jones, S.M., Klein, M., Weigle, M.C., & Nelson, M.L. (2020). MementoEmbed and Raintale for Web Archive Storytelling. In Web Archiving and Digital Libraries Workshop (WADL 2020). arXiv, abs/2008.00137. <https://arxiv.org/abs/2008.00137>

# With the DSA Toolkit, we can compare Archive-It collections about mass shootings



DSA Puddles Visualizations of the Dark and Stormy Archives Project

### Archive-It Collection 694: April 16 Archive

A collection curated by Virginia Tech: Crisis, Tragedy, and Recovery Network

A short summary story of 10 documents automatically selected by AlNoamany's Algorithm

« Previous Story      View the collection behind this story      Next Story »

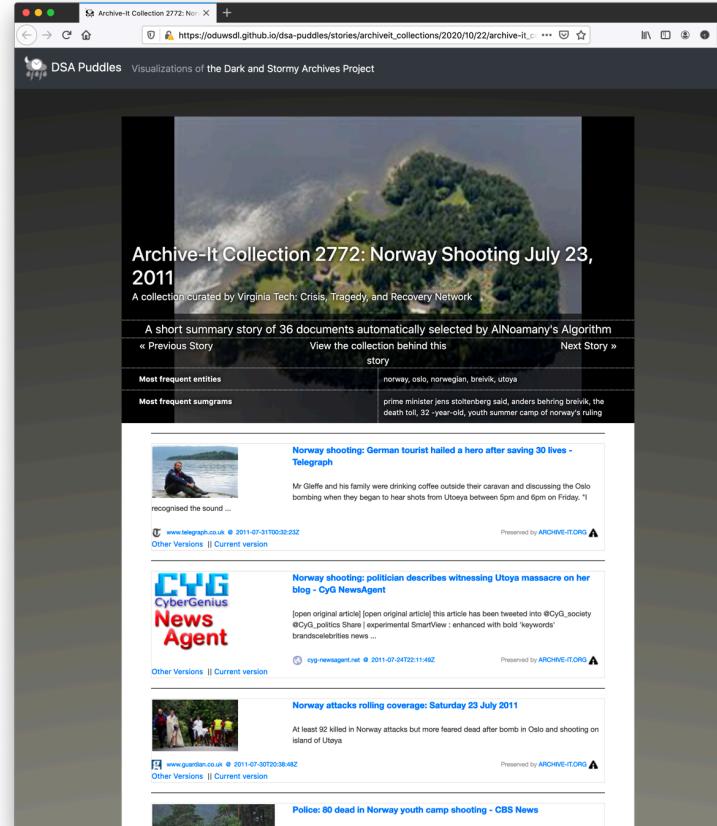
Most frequent entities: virginia tech, virginia, norris hall, blacksburg, the associated press

Most frequent sumgramps: virginia tech, the associated press, norris hall, 32 people, blacksburg va

**Virginia Tech Marks a Year After the Massacre - The Lede - Breaking News - New York Times Blog**  
A memorial wreath was placed on a candle guarded by the Virginia Tech Corp of Cadets on the campus of Virginia Tech in Blacksburg, Va. (Photo: Rob Carr/Associated Press) By this time...

**Virginia Tech Shooting - News - The New York Times**  
Monday's shooting at Virginia Tech was among the deadliest attacks at a school. Other notable rampages:

**Virginia Tech shooting victims are honored by mourners**  
Thursday, April 17, 2008 Virginia Tech shooting victims are honored by mourners Sue Lindsey /



DSA Puddles Visualizations of the Dark and Stormy Archives Project

### Archive-It Collection 2772: Norway Shooting July 23, 2011

A collection curated by Virginia Tech: Crisis, Tragedy, and Recovery Network

A short summary story of 36 documents automatically selected by AlNoamany's Algorithm

« Previous Story      View the collection behind this story      Next Story »

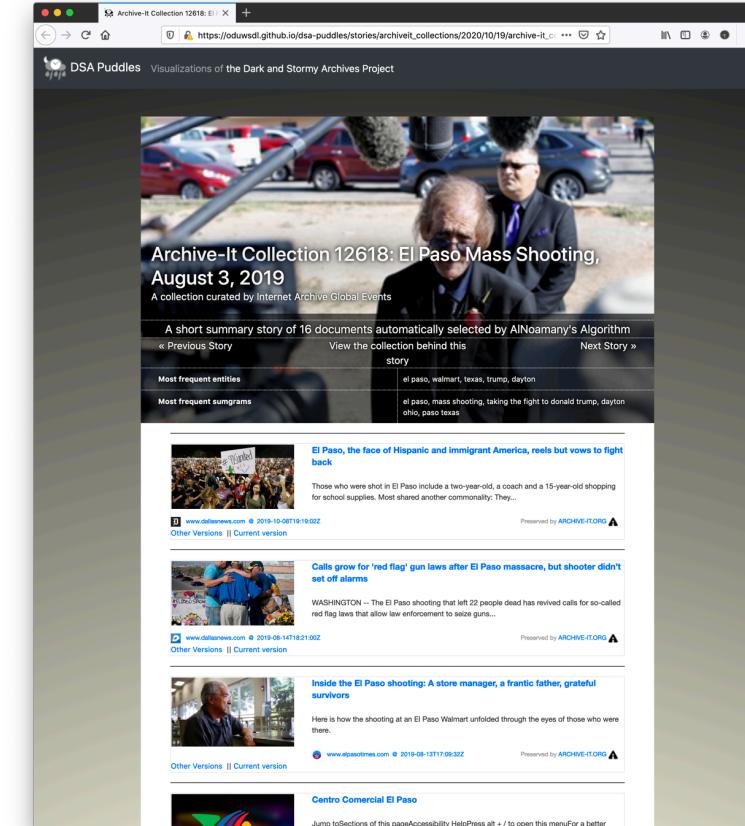
Most frequent entities: norway, oslo, norwegian, breivik, utoya

Most frequent sumgramps: prime minister jens stoltenberg said, anders behring breivik, the death toll, 32 year-old, youth summer camp of norway's ruling

**Norway shooting: German tourist hailed a hero after saving 30 lives - Telegraph**  
Mr Gieffe and his family were drinking coffee outside their caravan and discussing the Oslo bombing when they began to hear shots from Utøya between 8pm and 8pm on Friday. "I

**Norway shooting: politician describes witnessing Utøya massacre on her blog - CyG NewsAgent**  
[open original article] [open original article] this article has been tweeted into @CyG\_society @CyG\_politics Share | experimental SmartView : enhanced with bold keywords

**Norway attacks rolling coverage: Saturday 23 July 2011**  
At least 92 killed in Norway attacks but more feared dead after bomb in Oslo and shooting on island of Utøya



DSA Puddles Visualizations of the Dark and Stormy Archives Project

### Archive-It Collection 12618: El Paso Mass Shooting, August 3, 2019

A collection curated by Internet Archive Global Events

A short summary story of 16 documents automatically selected by AlNoamany's Algorithm

« Previous Story      View the collection behind this story      Next Story »

Most frequent entities: el paso, walmart, texas, trump, dayton

Most frequent sumgramps: el paso, mass shooting, taking the fight to donald trump, dayton ohio, paso texas

**El Paso, the face of Hispanic and immigrant America, reels but vows to fight back**  
Those who were shot in El Paso include a two-year-old, a coach and a 15-year-old shopping for school supplies. Most shared another commonality: They...

**Calls grow for 'red flag' gun laws after El Paso massacre, but shooter didn't set off alarm**  
WASHINGTON -- The El Paso shooting that left 22 people dead has revived calls for so-called red flag laws that allow law enforcement to seize guns...

**Inside the El Paso shooting: A store manager, a frantic father, grateful survivors**  
Here is how the shooting at an El Paso Walmart unfolded through the eyes of those who were there.

Virginia Tech



@shawnmjones @StormyArchives

Norway

45

Los Alamos National Laboratory

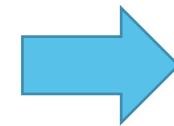
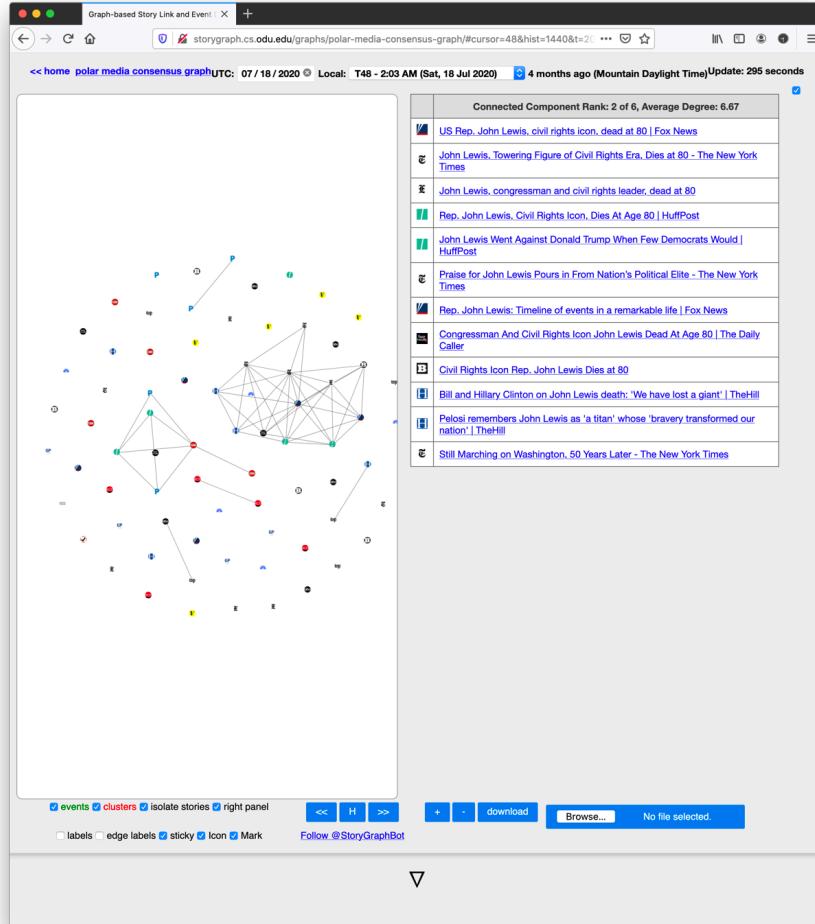


# But the DSA Toolkit is NOT just for Archive-It collections...

The DSA Toolkit works with Memento-compliant Archives



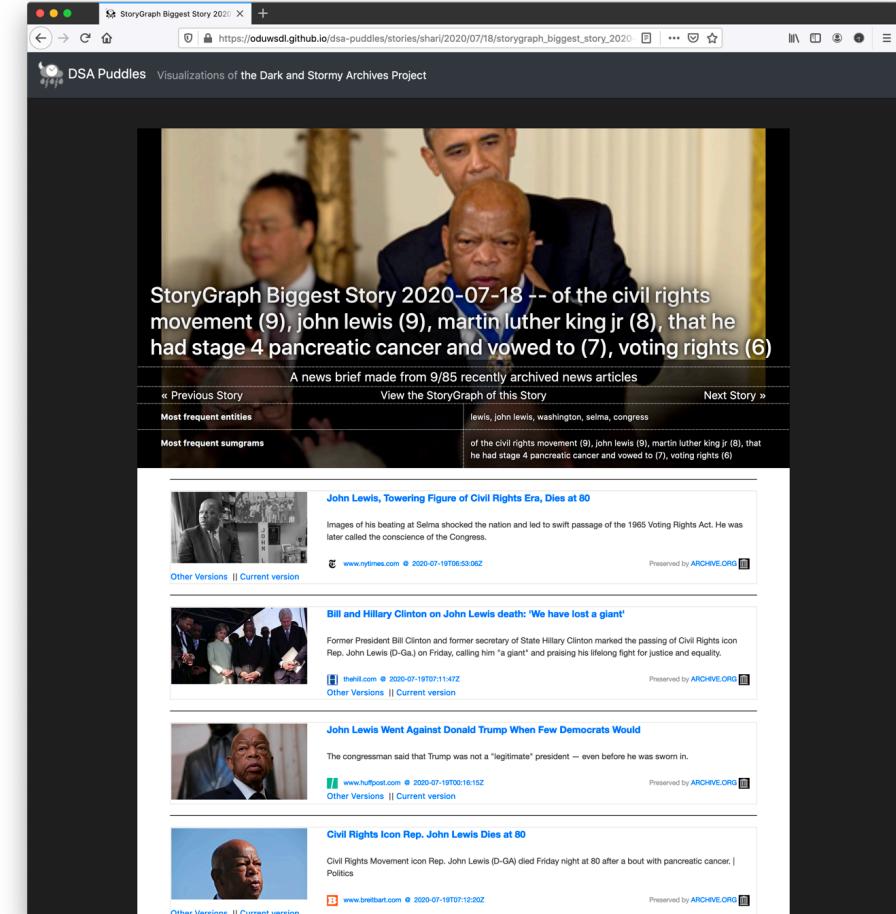
# We automatically produce daily news summaries



## MementoEmbed



## Raintale



To this: news articles of the biggest story of the day

From this: news articles extracted by StoryGraph



@shawnmjones @StormyArchives

Los Alamos National Laboratory



Jones, S.M., Nwala, A.C., Klein, M., Weigle, M.C., & Nelson, M.L. (2020). SHARI -- An Integration of Tools to Visualize the Story of the Day. In Web Archiving and Digital Libraries Workshop (WADL 2020). arXiv, abs/2008.00137. <https://arxiv.org/abs/2008.00137>

# This allows us to review news from the same day in different years

StoryGraph Biggest Story 2017-08-08 -- provocative action by launching ballistic missiles, nuclear weapons, north korea's nuclear, north korean, to miniaturize nuclear

A news brief made from 28/85 recently archived news articles

« Previous Story View the StoryGraph of this Story Next Story »

Most frequent entities: north korea, trump, us, the united states, pyongyang

Most frequent sumgrams: provocative action by launching ballistic missiles, nuclear weapons, north korea's nuclear, north korean, to miniaturize nuclear

North Korea to US: if you attack us, we'll respond with nukes

Trump promises North Korea 'fire and fury' over nuke threat

Trump threatens 'fire and fury' in response to North Korean threats

Trump: North Korea 'will be met with fire and fury like the world has never seen' if more threats emerge

August 8, 2017  
Nuclear Provocation by North Korea

StoryGraph Biggest Story 2018-08-08 -- special election, general election, ohio's 12 congressional district, republican troy balderson, close call

A news brief made from 28/85 recently archived news articles

« Previous Story View the StoryGraph of this Story Next Story »

Most frequent entities: republican, trump, democrats, democratic, ohio

Most frequent sumgrams: special election, general election, ohio's 12 congressional district, republican troy balderson, close call

Live results for the Washington primary elections

Live results for the 2016 Michigan primary elections

Live results for the Kansas primary elections

Live results for the Missouri primary elections

August 8, 2018  
US Primaries and Special Elections

StoryGraph Biggest Story 2019-08-08 -- el paso, dayton ohio, mass shootings, paso texas, august 7, 2019

A news brief made from 27/85 recently archived news articles

« Previous Story View the StoryGraph of this Story Next Story »

Most frequent entities: trump, el paso, dayton, donald trump, democratic

Most frequent sumgrams: el paso, dayton ohio, mass shootings, paso texas, august 7, 2019

Trump Is To Blame For Mass Shootings Because He Refused To Fight White Supremacist Domestic Terrorism

Puerto Rico just kicked out another governor

Puerto Rico Supreme Court Overturns Pedro Pierluisi As Governor

One of Trump's First Actions Reversed An Obama Rule On Guns And Mental Impairment

Beto O'Rourke Flatly Calls Trump A White Supremacist On Live TV

August 8, 2019  
Aftermath of Shootings at El Paso and Dayton

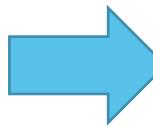


@shawnmjones @StormyArchives

# We can automatically summarize Wikipedia references



This image shows a screenshot of the Wikipedia article for Ruth Bader Ginsburg. The page includes her biography, a portrait photo, and a list of her professional roles and achievements, such as her service on the Supreme Court and her work as a lawyer.

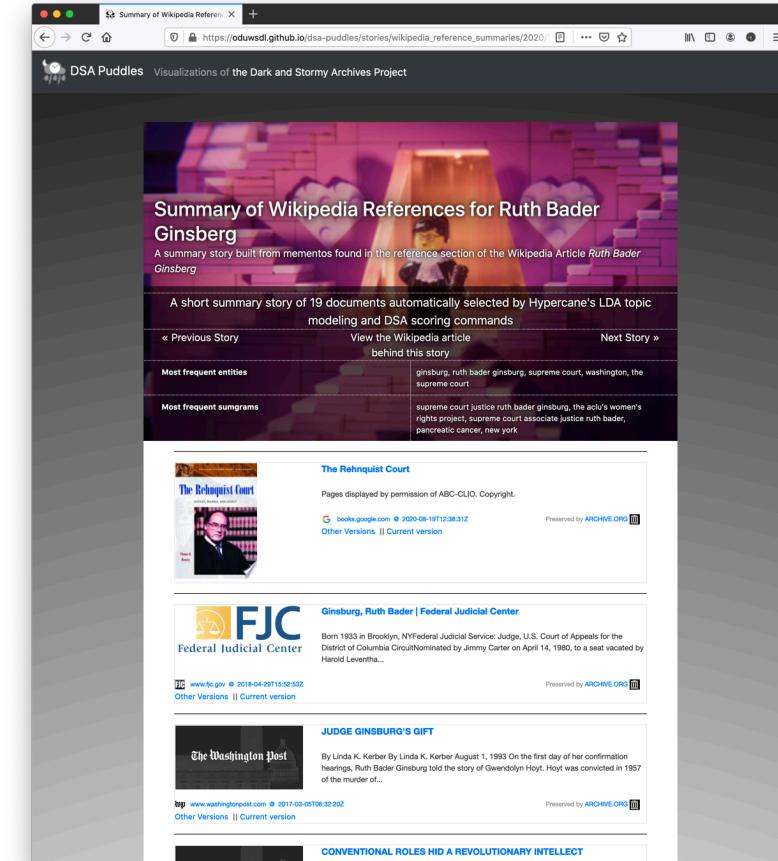


## MementoEmbed



## Raintale

From this: Wikipedia references as mementos



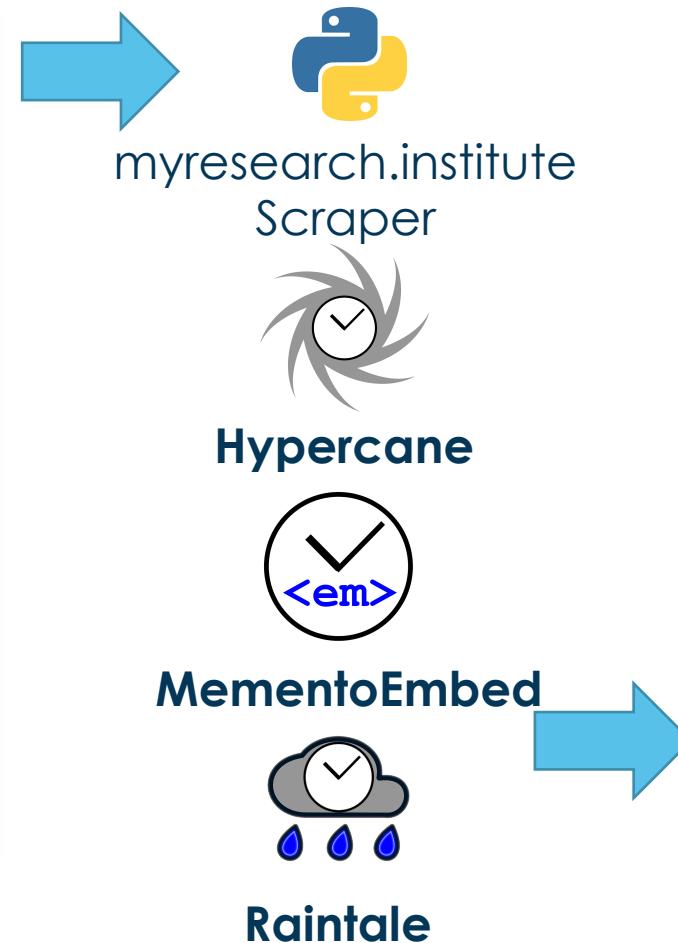
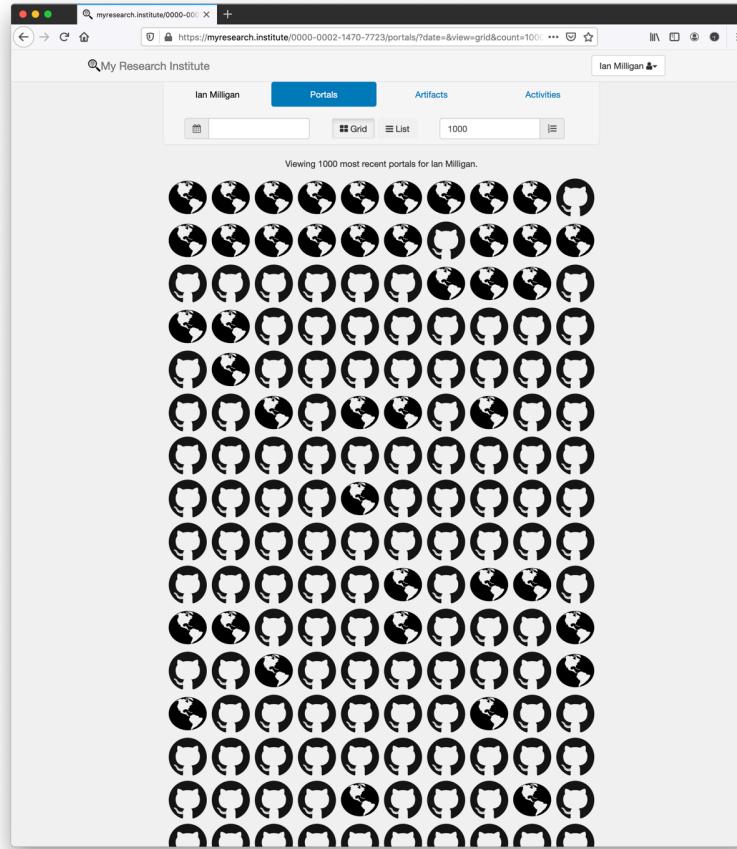
This image shows a screenshot of the DSA Puddles visualization for Ruth Bader Ginsburg. It displays a summary of 19 documents automatically selected by Hypercane's LDA topic modeling and DSA scoring commands. The visualization includes a thumbnail of a video, a summary text, and links to the original documents on Google Books and the Federal Judicial Center.

To this: a sample providing new insight into the references not seen in the article

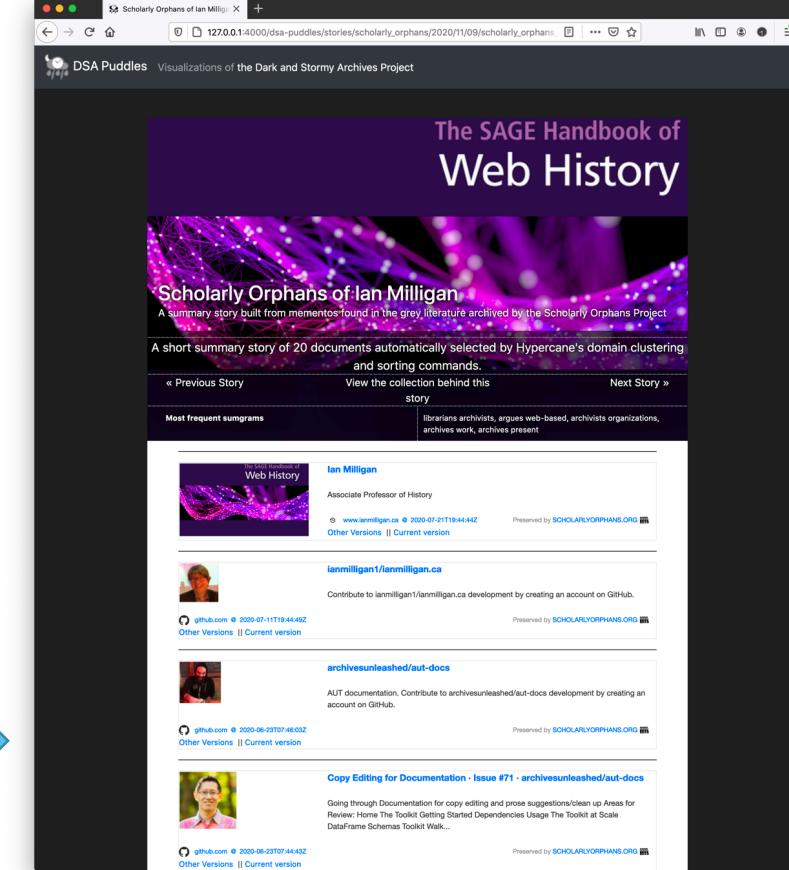


@shawnmjones @StormyArchives

# We can summarize a scholar's grey literature from the web archive collections of the Scholarly Orphans project, previously presented at CNI



From this: 1000 mementos of grey literature from the Scholarly Orphans project



To this: 20 mementos representing a scholar's work

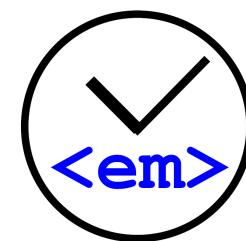
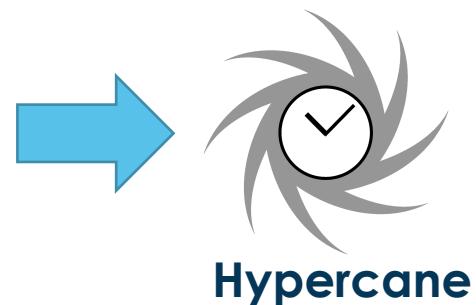


@shawnmjones @StormyArchives

Los Alamos National Laboratory



# You can automatically summarize your web archive collection here



From this: **your web archive collection**



A screenshot of a web interface showing a list of mementos for the "Occupy Movement 2011/2012" collection. The interface includes a header with "My Archive", "About", and "Contact" links, and a search bar. The main content area displays several cards, each representing a memento with a thumbnail, title, and a brief description. The cards include: "Occupy Camp Kinderland - Forward Thinking - Forward.com", "Occupy US | Mobile Broadcast News", "Police Use Flash Grenades To Reopen Seattle Terminal : NPR", "Occupy Sydney evicted from Martin Place on its 111th day of occupation. Media Release / Press Conference Information - OCCUPY SYDNEY | OCCUPY SYDNEY", and "May Day Special: The OccupyUSA Blog for May 1, With Frequent Updates! | The Nation".

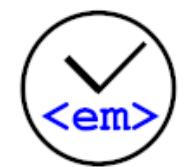
To this: a small sample of mementos visualized with your templates and features



@shawnmjones @StormyArchives

Jones, S.M., Klein, M., Weigle, M.C., & Nelson, M.L. (2020). MementoEmbed and Raintale for Web Archive Storytelling. In Web Archiving and Digital Libraries Workshop (WADL 2020). arXiv, abs/2008.00137. <https://arxiv.org/abs/2008.00137>

# For More Information on Dark and Stormy Archives



- Dark and Stormy Archives Project
  - Website: <https://oduwsdl.github.io/dsa/>
  - Example visualizations: <https://oduwsdl.github.io/dsa-puddles/>
- Hypercane – Intelligently Sample Mementos
  - Website: <https://oduwsdl.github.io/hypercane/>
  - Documentation: <https://hypercane.readthedocs.io/en/latest/>
  - Source code: <https://github.com/oduwsdl/hypercane>
- Raintale – Tell Stories with Many Mementos
  - Website: <https://oduwsdl.github.io/raintale/>
  - Documentation: <https://raintale.readthedocs.io/en/latest/>
  - Report Issues: <https://github.com/oduwsdl/raintale/issues>
  - Technical Report: <https://arxiv.org/abs/2008.00137>
- MementoEmbed – Summarize Individual Mementos
  - Documentation: <https://mementoembed.readthedocs.io/en/latest/>
  - Report Issues: <https://github.com/oduwsdl/MementoEmbed/issues>
  - Technical Report: <https://arxiv.org/abs/2008.00137>





# Backup slides



@shawnmjones @StormyArchives

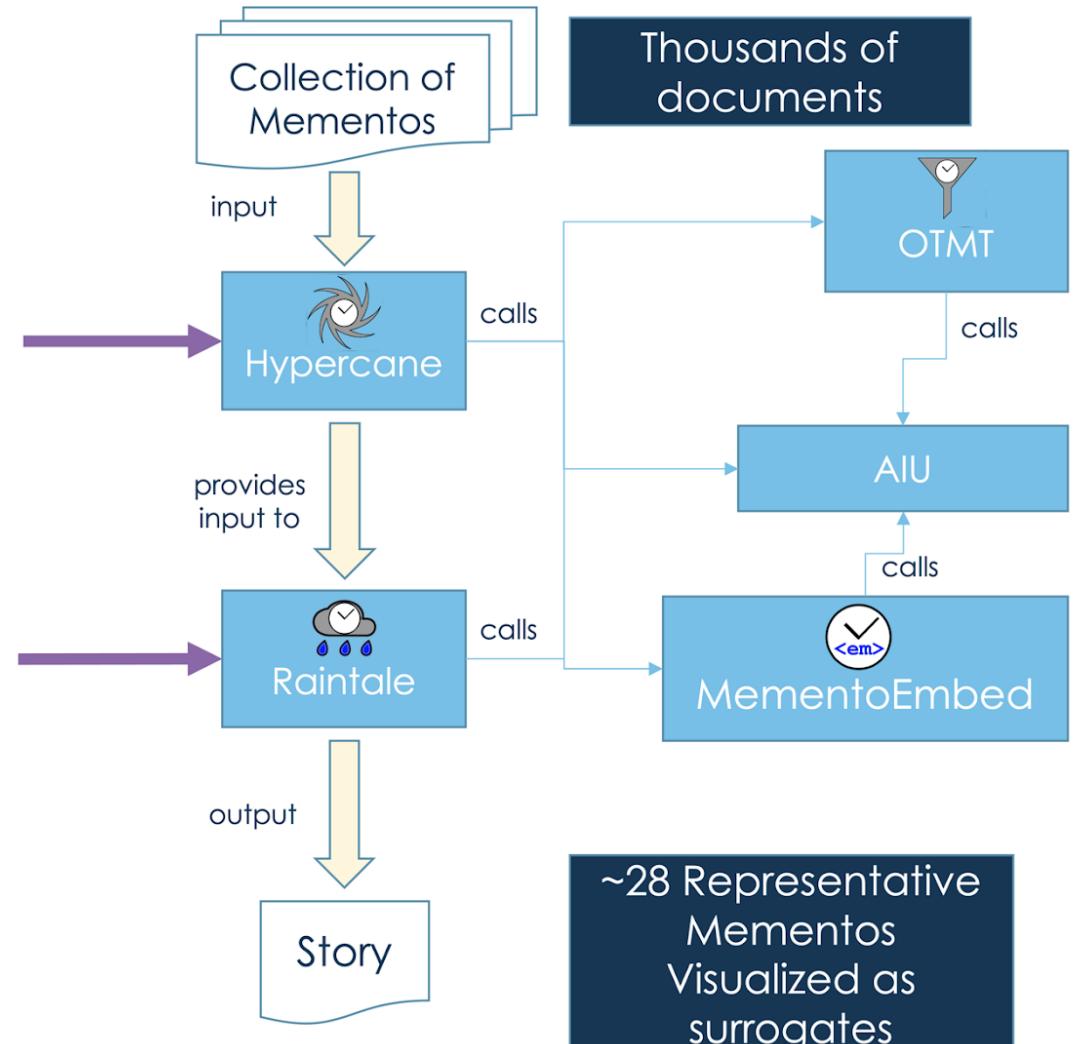
# The Dark and Stormy Archives Toolkit



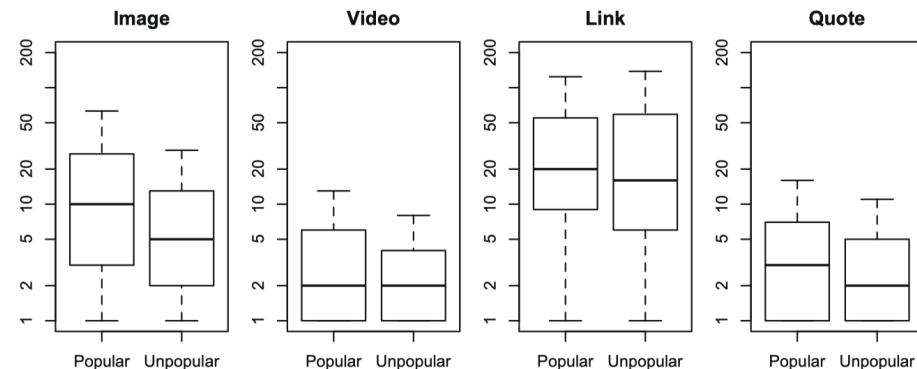
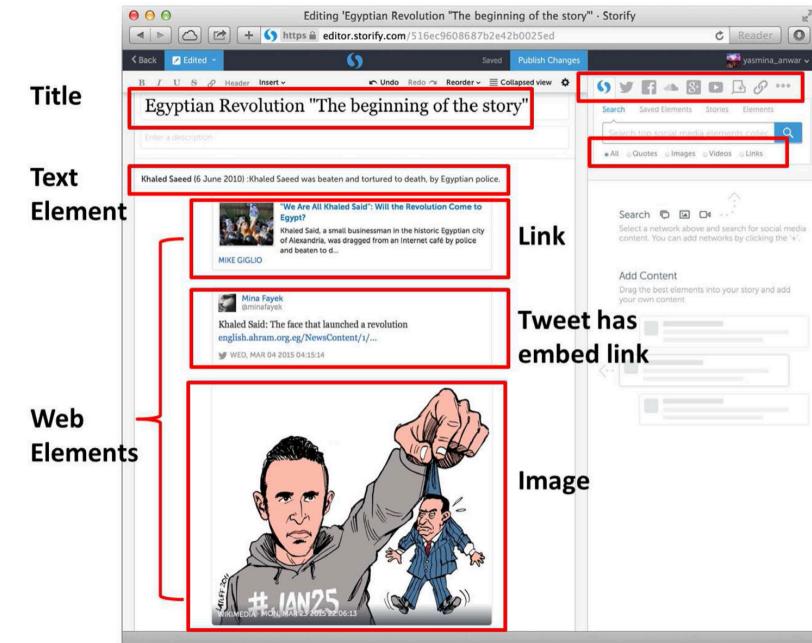
Tools for intelligently sampling mementos from a collection

Tools for visualizing mementos as a story

<https://oduwsdl.github.io/dsa/>



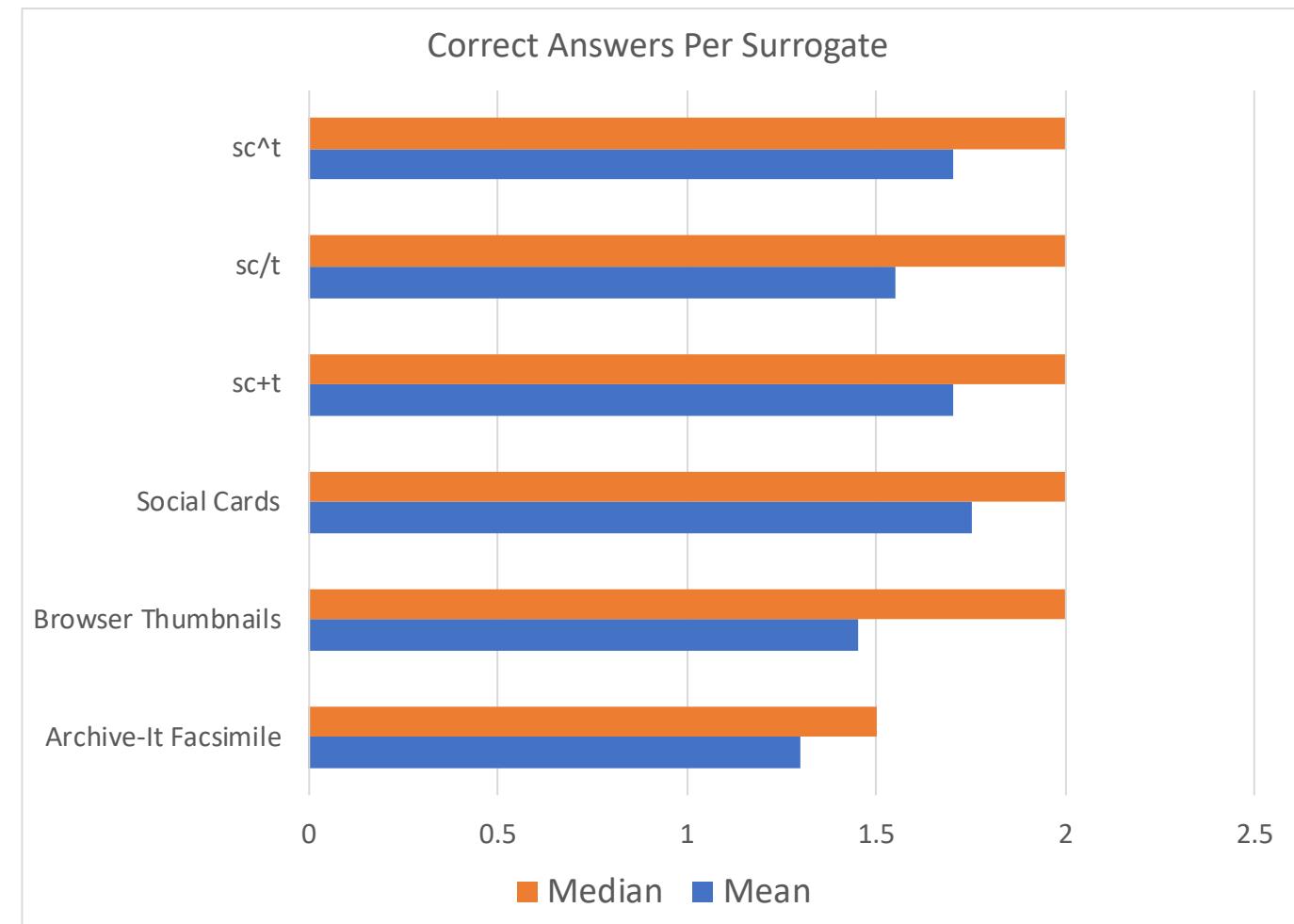
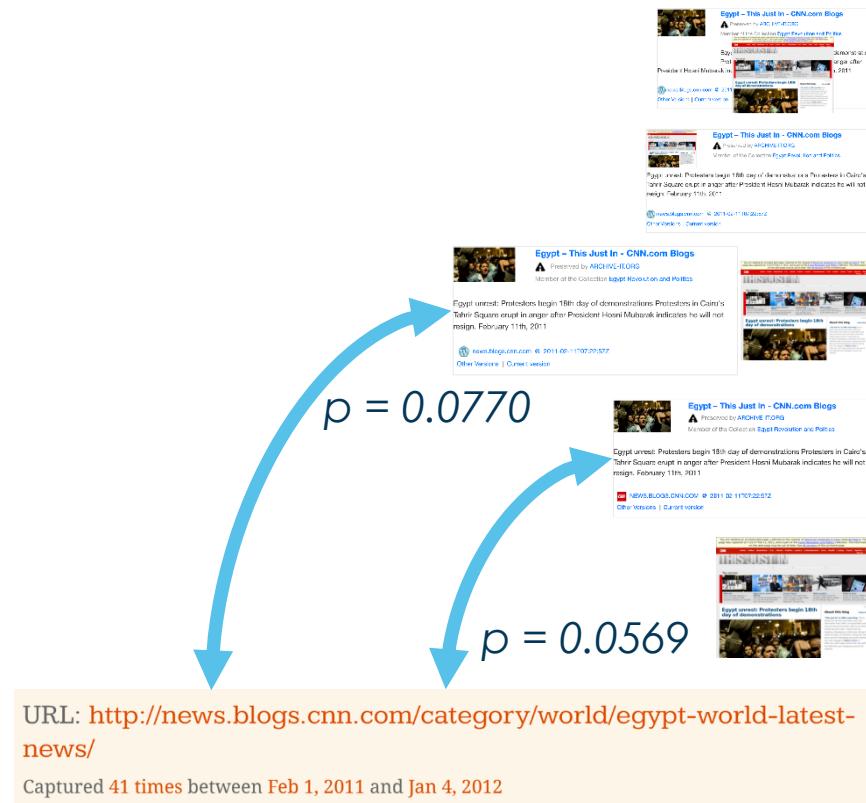
AI Noamany discovered that popular social media stories contain around 28 elements



@shawnmjones @StormyArchives

Y. AI Noamany, M. C. Weigle, and M. L. Nelson, "Characteristics of social media stories: What makes a good story?", International Journal on Digital Libraries, vol. 17, no. 3, pp. 239–256, 2016. <https://doi.org/10.1007/s00799-016-0185-3>

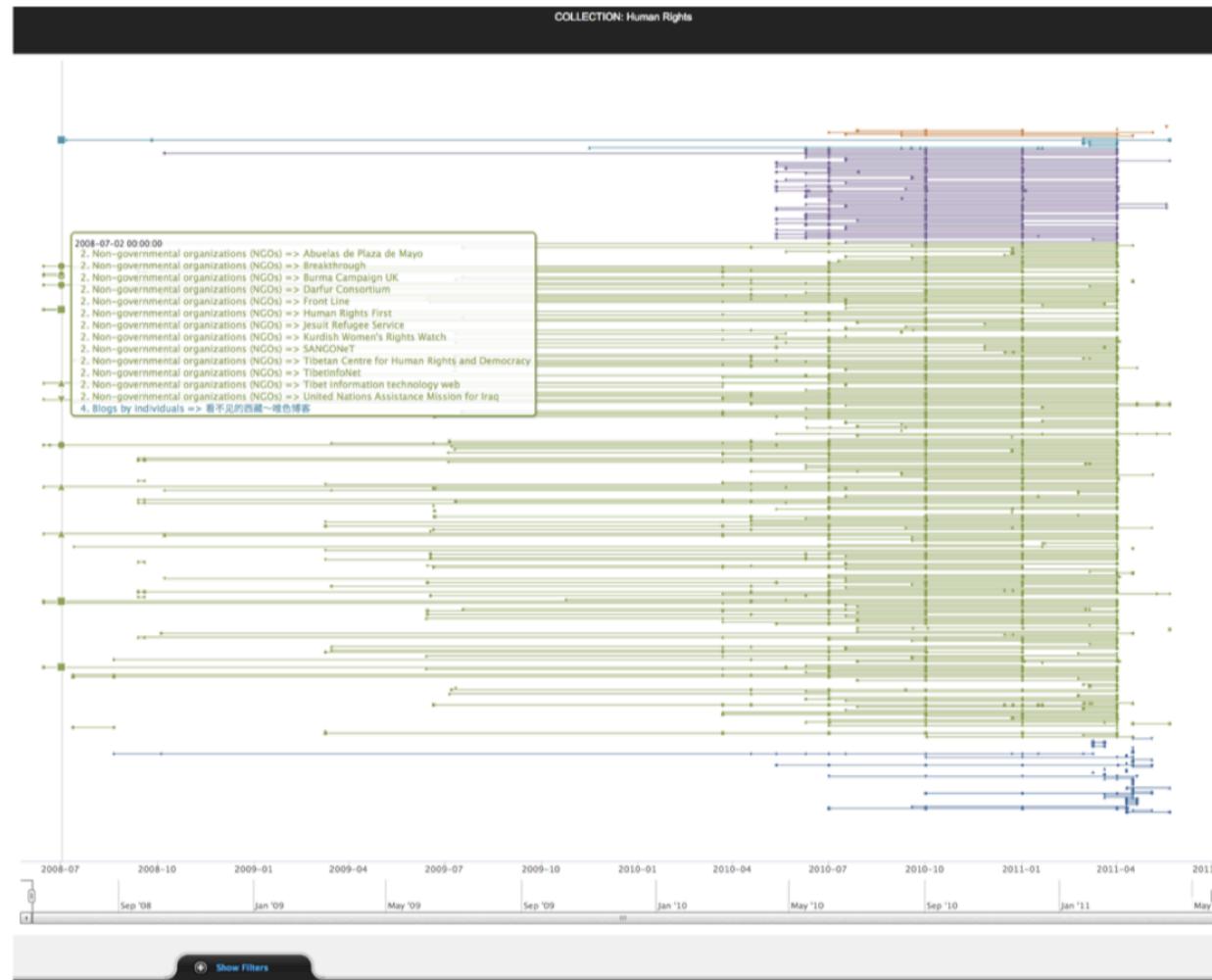
# Through a user study we discovered that social cards probably provide for better understanding of web archive collections



Jones, S.M., Weigle, M.C., & Nelson, M.L. (2019). Social Cards Probably Provide For Better Understanding Of Web Archive Collections. In Proceedings of the 28th ACM International Conference on Information and Knowledge Management (CIKM '19). <https://doi.org/10.1145/3357384.3358039>



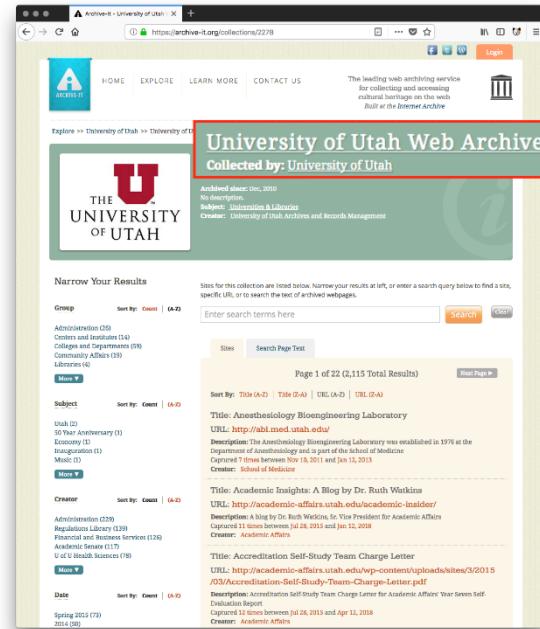
# Maybe timelines are best due to the temporal nature of web archives...



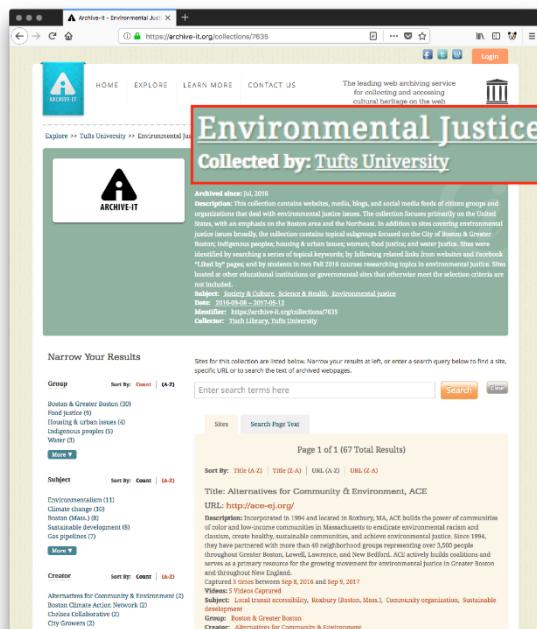
What do these colors mean?  
 If I hover over one, I get a group of titles?  
 How do I see across timelines?  
 What timeline do I pick first?  
 Which timelines are useful?  
 Why do so many stop at the same point?  
 Where are the seeds?  
 Where are the mementos?



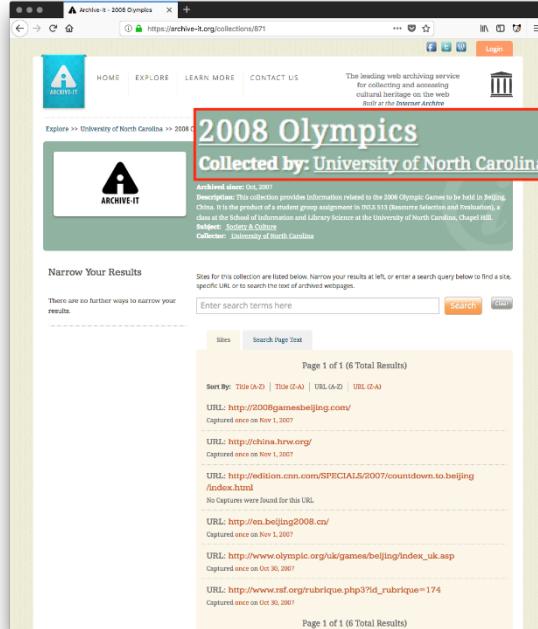
## We discovered four categories in Archive-It collections...



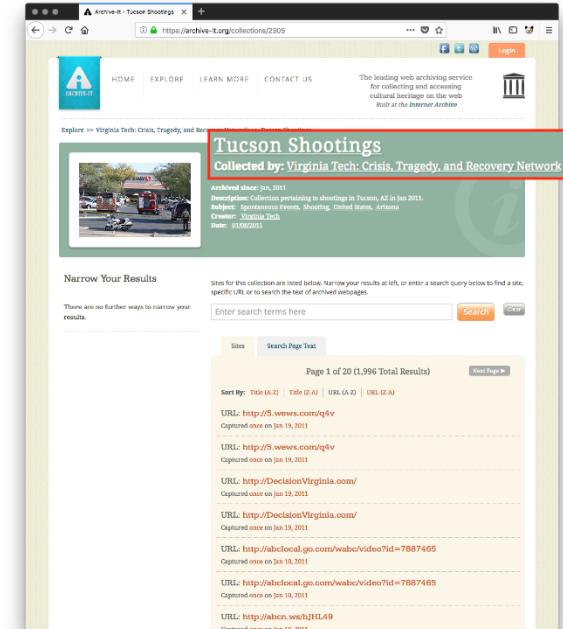
## Self-Archiving 54.1% of collections



Subject-based  
27.6% of collections



Time Bounded – Expected  
14.1% of collections



Time Bounded – Spontaneous  
4.2% of collections

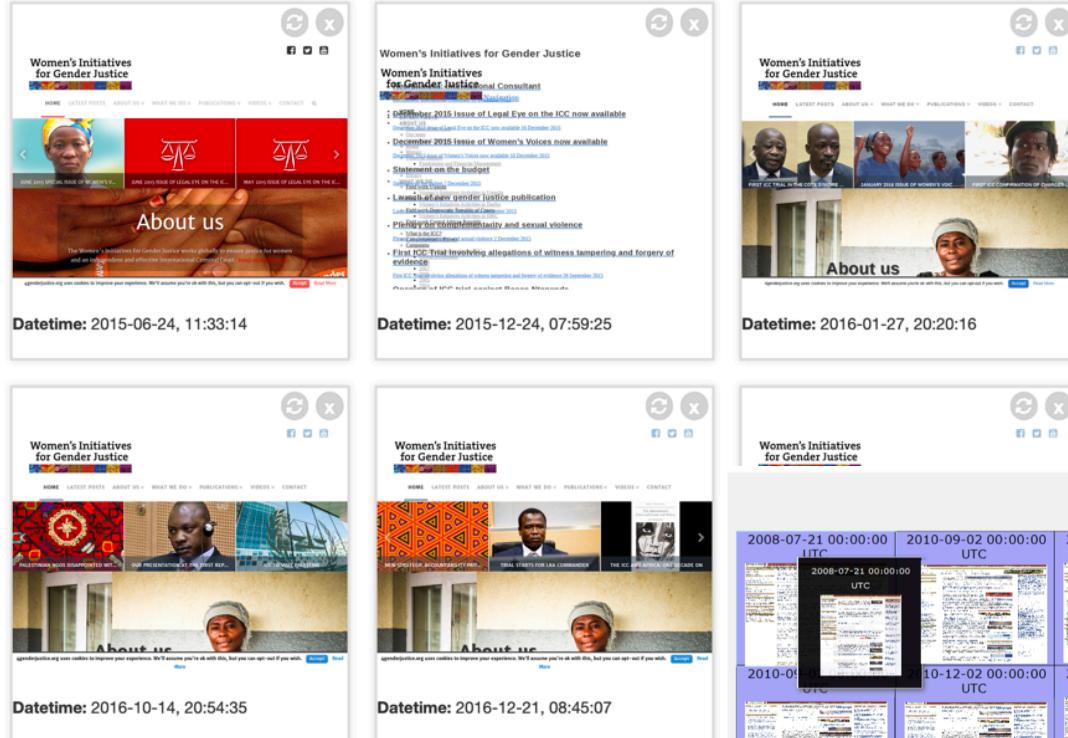
In a study of 3,382 Archive-It collections



@shawnmjones @StormyArchives

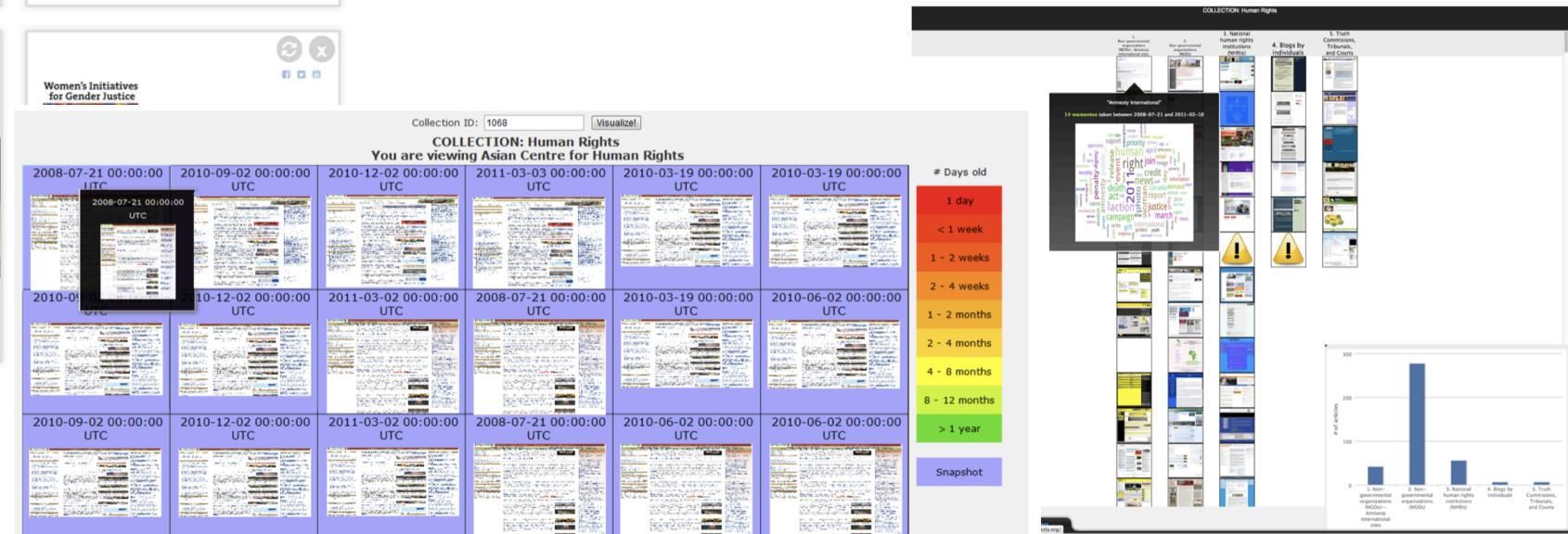
S. M. Jones, A. Nwala, M. C. Weigle, and M. L. Nelson. 2018. The Many Shapes of Archive-It. In *International Conference on Digital Preservation (iPRES) 2018*. <https://doi.org/10.17605/OSF.IO/EV42P>

# Perhaps thumbnails convey meaning...



With screenshots, we can view the evolution of a single web page, but what does it mean to view thumbnails across 1000s of pages?

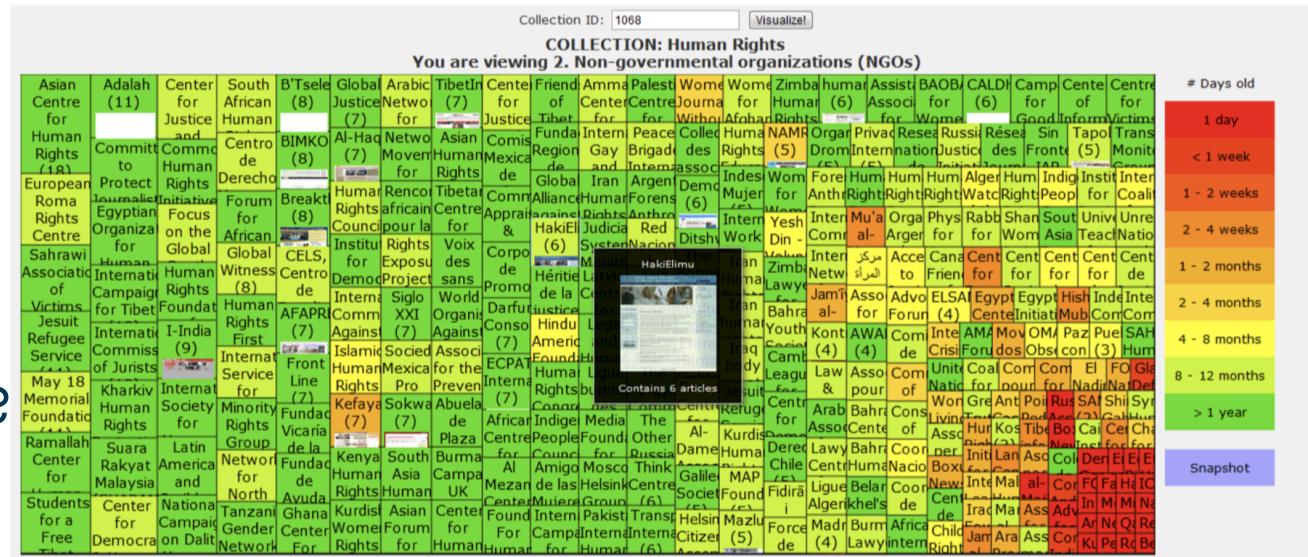
Thousands of mementos of different pages?



# Maybe heatmaps with colors are best for conveying what this collection is about...



What do the colors mean?  
Why are the differently sized blocks?  
Do I need to read the thumbnail?

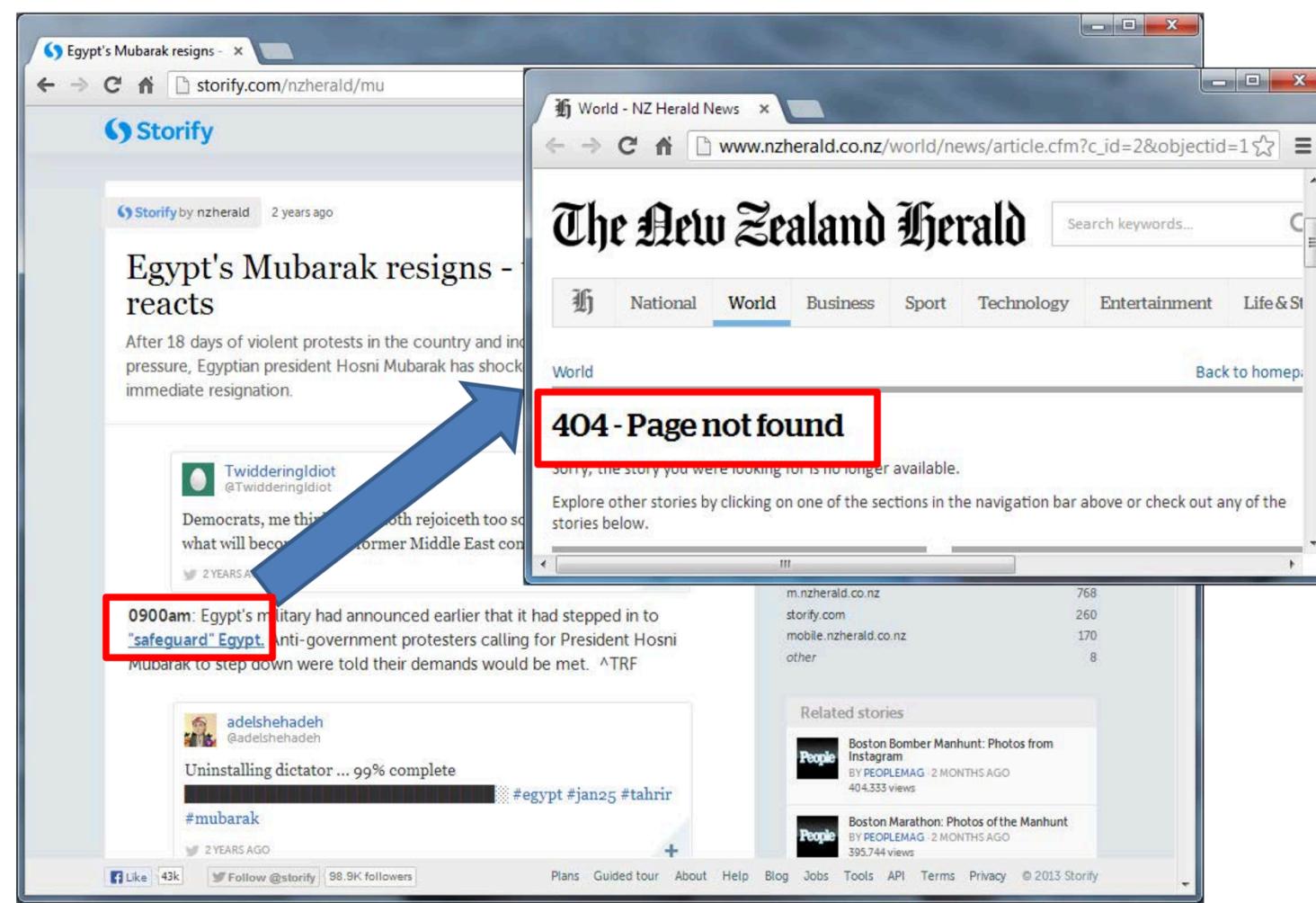


Are these times based on today?  
Is this the whole collection?  
Do I need to explore each block?



@shawnmjones @StormyArchives

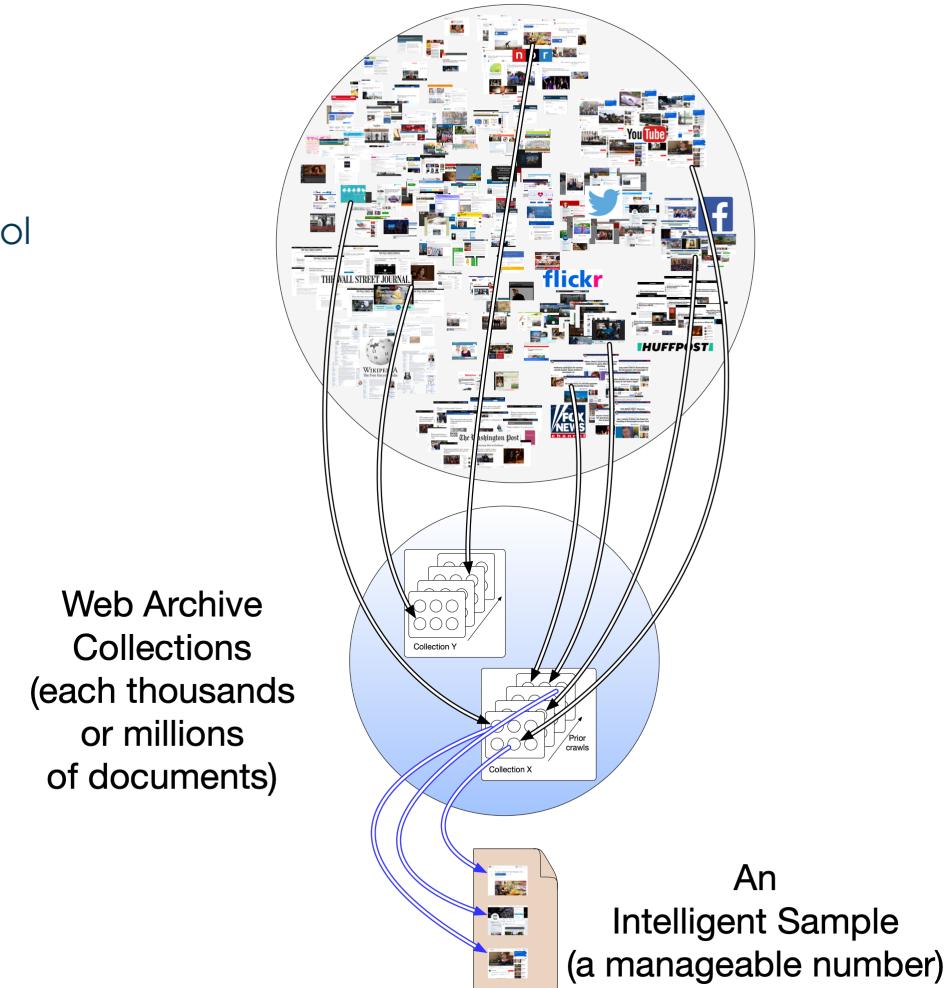
# Existing storytelling services are bookmarking, not preserving!



# Hypercane provides intelligent sampling of web archive collections

- **Works with Memento-compliant archives**
- Discovers mementos from TimeMaps and vice versa
- Creates new mementos with the help of ArchiveNow
- Can re-synthesize parts of web archives via the Memento Protocol
- Reports on:
  - collection metadata
  - named entities
  - terms
  - image analysis
- Synthesizing mementos into WARCs or other formats
- Clustering mementos by different algorithms
- Filtering mementos by off-topic, language, keyword matching, and more
- Scoring mementos by different functions
- Ordering mementos by features
- Currently a command-line application

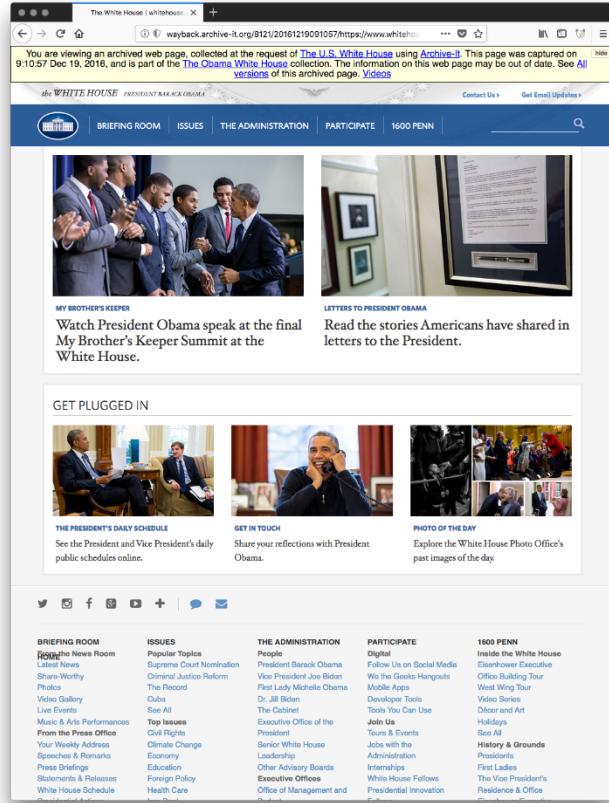
The Web in the “perpetual now”  
(near  $\infty$  documents)



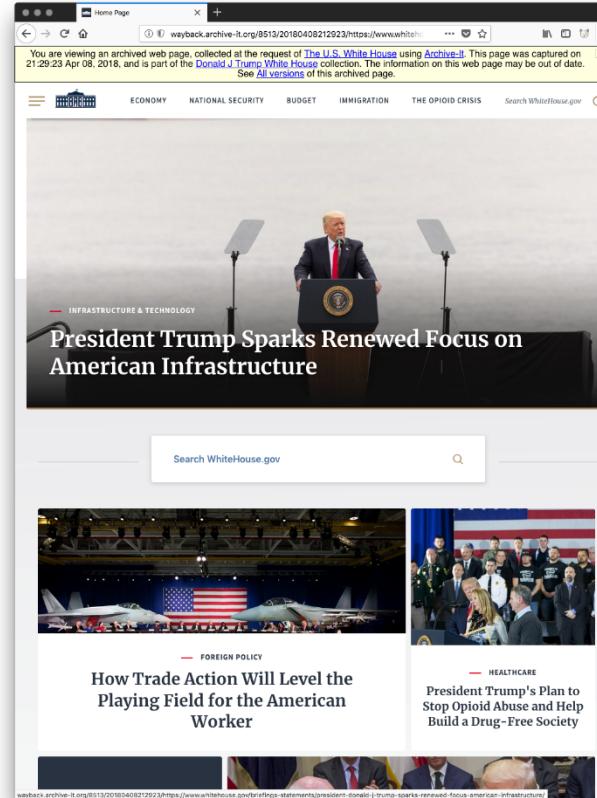
@shawnmjones @StormyArchives

S. M. Jones. “Hypercane Part 1: Intelligent Sampling of Web Archive Collections” <https://ws-dl.blogspot.com/2020/06/2020-06-03-hypercane-part-1-intelligent.html>, 2020.

# Different versions allow us to see changes in an organization's web presence



The White House: 2016



The White House: 2018



# Storytelling in social media consists of sampling and arranging web resources for summarization

Boston Marathon Bombing 2013

by alt\_stories 8 months ago 60 Views

- Watertown Residents Hunker Down As Manhunt Unfolds : NPR
- Search for marathon bombing suspect locks down Watertown, surrounding communities
- 19 Apr 2013 - Boston Marathon Explosions: Thursday's Developments : The Two-Way : NPR
- President Obama: "The American People Refuse to be Terrorized" | The White House
- 18 Apr 2013 - Obamas join Boston at memorial for marathon bombing victims
- Reddit and 4chan Investigate Boston Marathon Attack

Archive-It Stories

Boston Marathon Bombing 2013

29 Comments

- Watertown Residents Hunker Down As Manhunt Unfolds
- Search for marathon bombing suspect locks down Watertown, surrounding...
- Boston Marathon Explosions: Thursday's Developments
- President Obama: "The American People Refuse to be Terrorized"
- Obamas join Boston at memorial for marathon bombing victims
- The Internet Mystery-Solving Machine

Archive-It Stories

POSTS LIKES FOLLOWING ARCHIVE

Watertown Residents Hunker Down As Manhunt Unfolds : NPR

While the manhunt for a suspect in the Boston Marathon bombings continued Thursday night into Friday morning, residents of Watertown and surrounding communities were hiding in bedrooms, looking out from roofs and peering from behind locked doors. Many did not sleep as helicopters swirled overhead.

Search for marathon bombing suspect locks down Watertown, surrounding...

With nearly a million residents of the Boston area hunkering down behind locked doors, heavily armed police are making their way through a 20-block area of Watertown, searching for one of the suspects in the deadly Boston Marathon terror bomb attacks, the head of the State Police said this afternoon.



@shawnmjones @StormyArchives