# Curating the Internet: The Collaborative Future of Library Collections

Fiesole Collection Development Retreat Lille, April 20, 2017 Michael Levine-Clark University of Denver

# Expanding the Collection

# We concentrate our efforts on a very small portion of a huge potential collection

The Universe of Information

Scholarly Resources

Traditional Collection

# Expanding the Collection

# We concentrate our efforts on a very small portion of a huge potential collection

The Universe of Information

Scholarly Resources

Traditional Collection

# Expanding the Collection

# We concentrate our efforts on a very small portion of a huge potential collection

The Universe of Information Scholarly Resources Traditional Collection



The Universe of Information

 $\star$ 





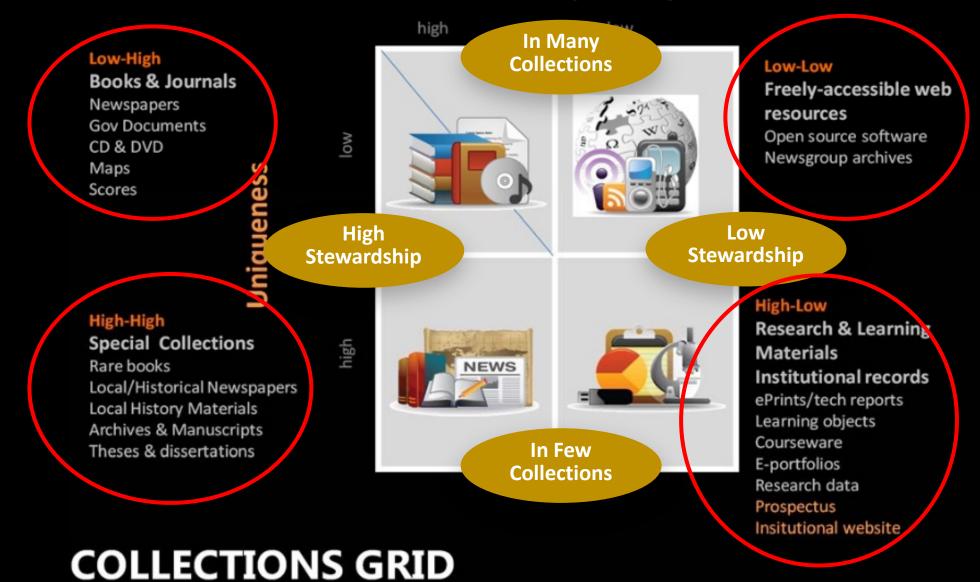
PORTICO

Your content. Preserved here.

PORTICO SERVICES & BENEFITS	ARCHIVE CONTENT & ACCESS	WHO PARTICIPATES IN PORTICO?	JOIN PORTICO	NEWS & EVENTS	ABOUT PORTICO
Addin	nillion file g more co e Archive? »	es preserve ontent ever	d. y day.		



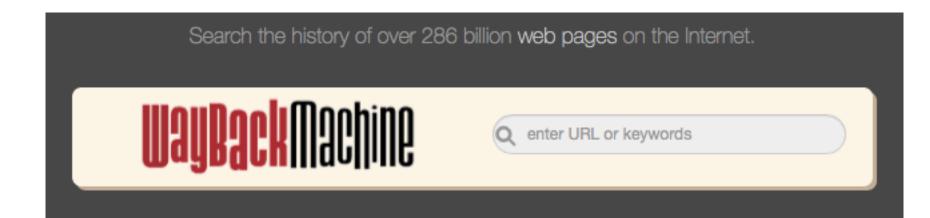
# Stewardship/scarcity



(Lorcan Dempsey and Eric Childress, OCLC Research)

**Internet Archive** is a non-profit library of millions of free books, movies, software, music, websites, and more.





https://archive.org/index.php

# **About the Internet Archive**

The Internet Archive is a 501(c)(3) non-profit library. Founded in 1996, our mission is to provide Universal Access to All Knowledge. We collect published works and make them available in digital formats. We are building a public library that can serve anyone in the world with access to the Internet.

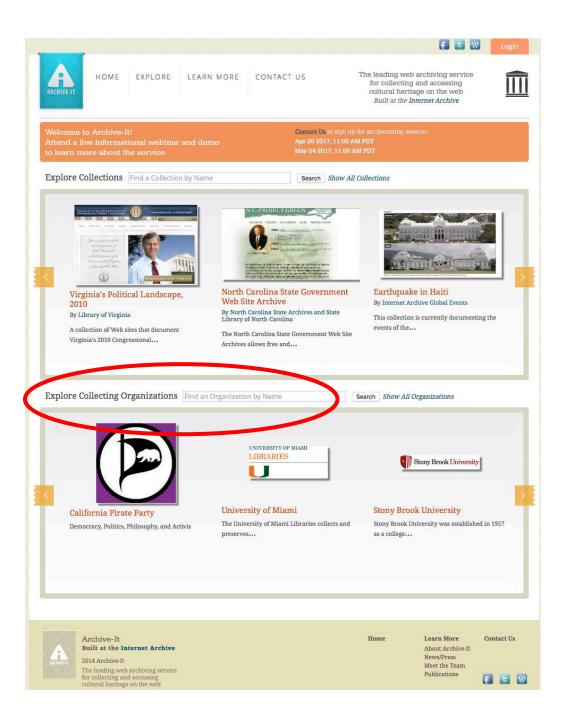
We began in 1996 by archiving the Internet itself, a medium that was just beginning to grow in use. Like newspapers, the content published on the web was ephemeral - but unlike newspapers, no one was saving it. Today we have 20+ years of web history accessible through the Wayback Machine and we work with 450+ library and other partners through our Archive-It program to identify important web pages.

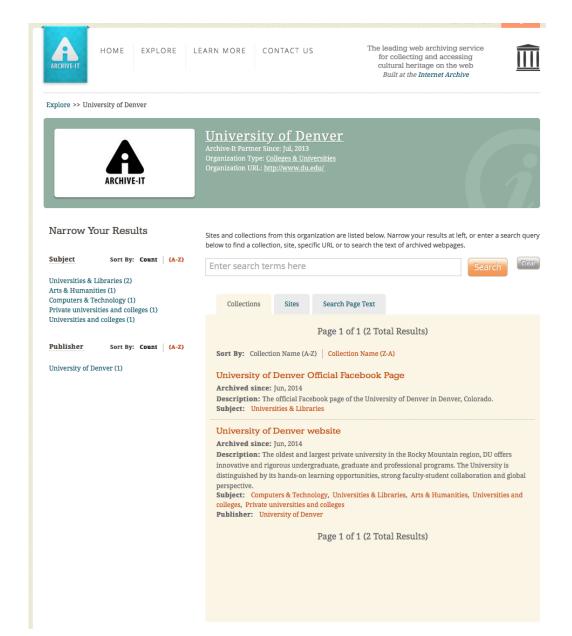


As our web archive grew, so did our commitment to providing digital versions of other published works. Today our archive contains:

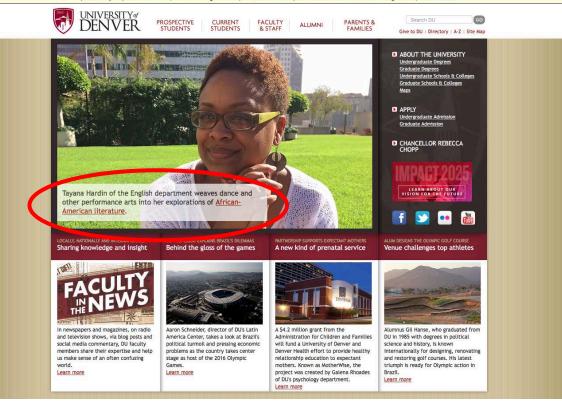
- 279 billion web pages
- 11 million books and texts
- 4 million audio recordings (including 160,000 live concerts)
- 3 million videos (including 1 million Television News programs)
- 1 million images
- 100,000 software programs

Anyone with a free account can upload media to the Internet Archive. We work with thousands of partners globally to save copies of their work into special collections.





You are viewing an archived web page, collected at the request of <u>University of Denver</u> using <u>Archive-It</u>. This page was captured on 21:59:07 Aug 08, 2016, and is part of the <u>University of Denver</u> website collection. The information on this web page may be out of date. See <u>All versions</u> of this archived page. Note that this document was downloaded, and not saved because it was a duplicate of a previously captured version (21:15:26 Aug 8, 2016). HTTP headers presented here are from the original capture. <u>Videos</u>





#### University of Denver website Web Archive (University of Denver)



All 📀 Take Me Back

#### Not in Archive

The page you requested has not been archived in Archive-It.

This could be for a number of reasons.

Most likely the page you are requesting was outside of the crawler's scope. Try another request or click to see other pages from news.du.edu.

Enter Web Address: http://

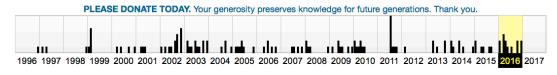
It is also possible that this site is currently being crawled and the archived pages are not yet available in the Wayback Machine. It can take up to 24 hours after a crawl has been completed for the site to appear in the Wayback Machine. Please try again after the allotted time.

You can also try searching for news.du.edu/new-partnership-with-denver-health-will-help-expectant-mothers/ on the live web, in the global Wayback Service, or in the General Archive at the Internet Archive at www.archive.org



Explore more than 284 billion web pages saved over time

#### Saved **122 times** between December 19, 1996 and December 25, 2016. Summary of gabf.org



			JAN	ı.						FEB	3					I	MAF	2						APF	2		
					1	2		1	2	3	4	5	6			1	2	3	4	5						1	2
3	4	5	6	7	8	9	7	8	9	10	11	12	13	6	7	8	9	10	11	12	3	4	5	6	7	8	9
10	11	12	13	14	15	16	14	15	16	17	18	19	20	13	14	15	16	17	18	19	10	11	12	13	14	15	16
17	18	19	20	21	22	23	21	22	23	24	25	26	27	20	21	22	23	24	25	26	17	18	19	20	21	22	23
24	25	26	27	28	29	30	28	29						27	28	29	30	31			24	25	26	27	28	29	30
31																											
		I	MA	(						JUN							JUL							AUG	•		
	2				6	7						3	4				JUL		1	2		1				5	6
	2		4	5			5	6			2			3	4					_	7	1 8	2	3	4	5 12	
1	2	<b>3</b> 10	4 11	5 12	13	14	-	-	7	1	2 9	10	11	-	4 11	5	6	7	8	9	-	8	2 9	3 10	4 11		13
1 8 15	2 9	<b>3</b> 10 17	4 11 18	5 12 19	13 20	14 21	12	13	7 14	1 8	2 9 16	10 17	11 18	10		5 12	6 13	7 14	8 15	9 16	14	8 15	2 9 16	3 10 17	4 11 18	12	13 20
1 8 15 22	2 9 16	<b>3</b> 10 17 24	4 11 18	5 12 19	13 20	14 21	12 19	13 20	7 14 21	1 8 15	2 9 16 23	10 17 24	11 18	10 17	11	5 12 19	6 13 20	7 14 21	8 15 22	9 16 23	14	8 15 22	2 9 16 23	3 10 17 24	4 11 18	12 19	13 20

															31													
			;	SEF	•					(	OCI	Г						101	/					I	DEC	;		
					1	2	3							1			1	2	3	4	5					1	2	3
4	1	5	6	7	8	9	10	2	3	4	5	6	7	8	6	7	8	9	10	11	12	4	5	6	7	8	9	10
1	1	12	13	14	15	16	17	9	10	11	12	13	14	15	13	14	15	16	17	18	19	11	12	13	14	15	16	17
1	8	19	20	21	22	23	24	16	17	18	19	20	21	22	20	21	22	23	24	25	26	18	19	20	21	22	23	24
2	5 2	26	27	28	29	30		23	24	25	26	27	28	29	27	28	29	30				25	26	27	28	29	30	31
								30	31																			

#### Note

This calendar view maps the number of times www.gabf.org was crawled by the Wayback Machine, *not* how many times the site was actually updated. More info in the FAQ. Green indicates redirects (3xx).

# WayBack Machine

## https://www.greatamericanbeerfestival.com/news/whats-r



## Hrm.

Wayback Machine doesn't have that page archived. Want to search for all archived pages under https://www.greatamericanbeerfestival.com/ ?

## This page is available on the web!

Help make the Wayback Machine more complete! Save this url in the Wayback Machine



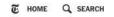
Search this blog GO

Print Subscribe Share/Save

About This Blog

# How Tweet It Is!: Library Acquires Entire Twitter Archive

April 14, 2010 by Matt Raymond



## The New York Eimes



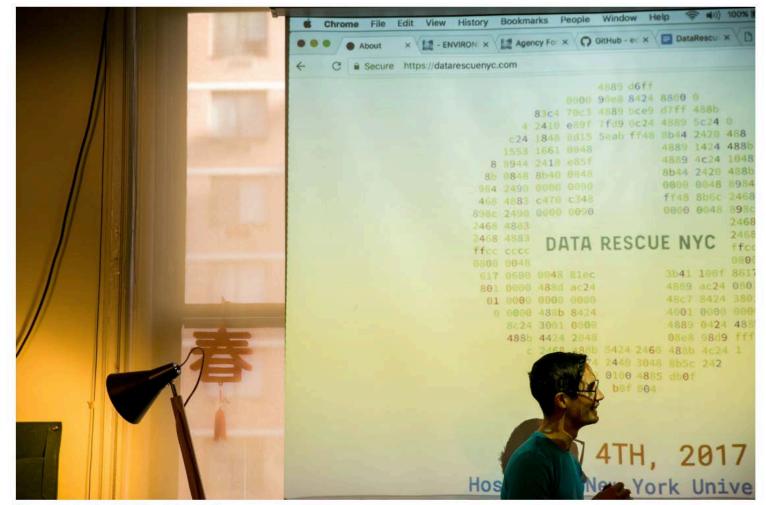


SCIENCE

# Activists Rush to Save Government Science Data — If They Can Find It

By AMY HARMON MARCH 6, 2017

## 



"Data rescuers" at New York University last month. Sam Hodgson for The New York Times

# DataRefuge

Search

# Q

# Building refuge for federal climate & environmental data

DataRefuge helps to build refuge for federal data and supports climate and environmental research and advocacy. We are committed to fact-based arguments. DataRefuge preserves the facts we need at a time of ongoing climate change.

This site is one part of the project. The vast majority of the government information gathered through this project is available from the Internet Archive through the End of Term project. This data catalog is a place to store data that is difficult or impossible to harvest through web crawlers. Beginning to add difficult data to this websites is one of the tasks of the DataRescuePhilly event on January 13th and 14th.



INTERNATIONAL INTERNET PRESERVATION CONSORTIUM

## A GLOBAL NETWORK OF EXPERTS ARCHIVING THE WEB FOR FUTURE GENERATIONS.

Learn More About IIPC







#### WHY ARCHIVE THE WEB?

Information published on the World Wide Web today will be the primary resources for future researchers.

Why Archive the Web?



# Dépôt légal des sites web

- Types de documents
- Modalités pour les éditeurs, producteurs et hébergeurs
- Cas des livres numériques (e-books)
- Attribution d'un ISSN ou d'un ISBN
- ISSN
- ISBN

## Types de documents

Le dépôt légal concerne tous les types de publications diffusées sur l'internet : sites institutionnels ou personnels, revues d'accès gratuit ou payant, blogs, sites commerciaux, plateformes de vidéos ou de livres numériques...

Les contenus édités ou les éditeurs doivent avoir un lien avec le territoire français. Il s'agit avant tout des sites web enregistrés en .fr ainsi que ceux des départements et territoires d'outre-mer. D'autres sites web en .com, .org, .net, etc. édités par des personnes physiques ou morales domiciliées en France font également partie du périmètre. Enfin, les publications produites sur le territoire national peuvent être concernées même si elles sont diffusées par une société étrangère.

Afin d'identifier ces sites web, la BnF a passé des accords avec des organismes en charge de la gestion de noms de domaines, comme l'Association française pour le nommage internet en coopération (AFNIC), la société de conseil en systèmes et logiciels informatiques (OVH) et l'Office des postes et télécommunications de Nouvelle-Calédonie (OPT-NC). Chaque année, plusieurs millions de sites sont ainsi capturés, jusqu'à une profondeur déterminée à l'avance (de l'ordre de quelques milliers de fichiers par site).



### Web Archiving at the Library of Congress

In 2000, the Library of Congress established a pilot project to collect and preserve websites. A multidisciplinary team of Library staff studied methods to evaluate, select, collect, catalog, provide access to, and preserve these materials for future generations of researchers.

Since then, the Library has developed thematic web and event-based archives on such topics as the United States National Elections, the Iraq War, and the events of September 11. Designated staff in Library Services and the Law Library select content for the archives.

A Web Archiving Team, currently located in the Library's Technology Policy office, supports the goal of managing and sustaining atrisk digital content. The team is charged with building a Library-wide understanding and technical infrastructure for capturing web content. The team, in collaboration with a variety of Library staff, and national and international partners, is identifying policy issues, establishing best practices and building tools to collect and preserve web content.

ibrary of Congress > Archived We	eb Sites		🔂 Subscribe 🛛 🖉 Share/Save			
IRMAT						
rchived W	Joh Sit	Δ				
					OF SS Archived Web Sites ~ Search	i Loc.gov
arch Archived Web Site	es Collecti	ons with Archived Web Sites				
				Library of Congress > Archived Web Sites	> Search	🖸 Subscribe 👩 Share/Sa
sults: 1-25 of 11,335   Refine	ed by: Original	Format: Archived Web Site 🗙		FORMAT	c'.	
6				Archived Web	Site	
fine your results		Search Archive	ed Web Sites	Search Archived Web Sites	Collections with Archived Web Sites	
Available Online	11,335		View List V Go Sort By Select V Go			
All Items	11,335		View List ~ Go Sort By Select ~ Go	Results: 1-25 of 2,310   Refined by:	Original Format: Archived Web Site X (Part of: Sep	tember 11, 2001 Web Archive \star
riginal Format		AUTSEAN-BURMA	COLLECTION	Refine your results	Search Archive	d Web Sites
chived Web Site	×	Manual Manual<	Burma/Myanmar General Election 2010 Web Archive		2,310	
line Format			Collection Period: May 2010 to January 2011. The November 7, 2010 Burma/Myanmar elections provided for under the 2008 constitution were watched by the world as	O All Items	2,310	View List ~ Go Sort By Select ~
eb Page	11,277		critical to a government seeking broader legitimacy in the road to democracy. Most	Original Format	A1	COLLECTION
-	27		governments, including the United States, criticized the elections as not meeting international standards of free and fair elections. Two main parties contested: the	Archived Web Site	× You Will Never	September 11, 2001 Web Archive
ine Text	1		Union Solidarity and	Online Format	Be Fragment Baseline 7-300 partner baset garbanette	Collection Period: September 11, 2001 to December 1, 2001. The September 11, 200 Web Archive preserves the web expressions of individuals, groups, the press and institutions in the lighted States and from acround the world in the affectments of the
dio	1	Name And Annual An Annual Annual Annu	Contributor: Library of Congress	Web Page PDF	3	institutions in the United States and from around the world in the aftermath of the attacks in the United States on September 11, 2001. The selected websites are comprised broadly of United States and non-United States government sites; press,
			Date: 2010 View 34 Items	Audio	1	Contributor: Library of Congress - Internet Archive (Firm) - Webarchivist.org
te	44.000			Date		(Organization) - Pew Internet & American Life Project Date: 2004
00 to 2099	11,330 1			2000 to 2099	2,310	View 2,311 Items
00 to 1999	1	ROCKWOOD	ARCHIVED WEB SITE Official Campaign Web Site - Christopher B. Rockwood	Location		
cation		FOR CONGRESS	Web Site. electronic   Electronic (Form).		1,702	ARCHIVED WEB SITE The Miami Herald
ited States	7,798	About Date: Insure: Volumer: New j. Generative Bouer Gantaction About Chris	Contributor: Rockwood, Christopher	Afghanistan United Kingdom	1,031 Che stillami Heraib	The Miami Herald - Latin America The Miami Herald, a Web Site produced by Miami Herald, a press organization, is pa of the Library of Congress September 11 Web Archive and preserves the web
ghanistan	1,052		Date: 2014-08-19	India		expressions of individuals, groups, the press and institutions in the United States an
lifornia	550			Germany	31	from around the world in the aftermath of the attacks in the United States on September 11, 2001. Web Site. electronic   Electronic
kas	394			Canada Pakistan	25 <b>24</b>	Contributor: Miami Herald Date: 2001-10-02
rida	313	Mate by 2000 fault Chen Ricchened was born in McRigan, where the grandfaller was a union thereof during an E-year career at Find Mater Scorpany, and your at more Philologic inite Mater, a Mater and water to the Study struct. Its indeer was againster by fair succession guments or the Perindpatas Cargoritem of		Japan	21	vate: 2001-10-02
w York	307	Environment Protecture's Educes Abilitary Example		Argentina	20	
chigan	251		ARCHIVED WEB SITE	France More Locations »	16	ARCHIVED WEB SITE Interpol
nois	220		Official Campaign Web Site - Shaun Lynn McCausland			Interpol Interpol, a Web Site produced by Interpol, an international intergovernmental body,
dan	219		Web Site. electronic   Electronic (Form).	Part of September 11, 2001 Web Archive		part of the Library of Congress September 11 Web Archive and preserves the web expressions of individuals, groups, the press and institutions in the United States an
nnsylvania	212	To enjoy this site you'll need to update your Planh Player. It's easy, painteen and will take just a moment	Contributor: McCausland, Shaun Lynn	Digital Collections		from around the world in the aftermath of the attacks in the United States on September 11, 2001. Web Site. electronic   Electronic (Form).
re Locations »		Deventing and install the latest version of Added's Plant Play, ver D. Minsteinanty you'r their meet on class your west braneau K. Door kurther late with your nester your west braneau K. Those Joints fort	Date: 2012-03-23	Humanities and Social Sciences Division	1 · · · · · · · · · · · · · · · · · · ·	Contributor: Interpol Date: 2001-09-24
rt of		Click to download		Contributor		Marce Evolation
imanities and Social	6,170			Contributor		
nited States Elections Web rchive	5,405					
eptember 11, 2001 Web	2,310	e e e e e e e	ARCHIVED WEB SITE			

https://www.loc.gov/websites/



Search Loc.gov

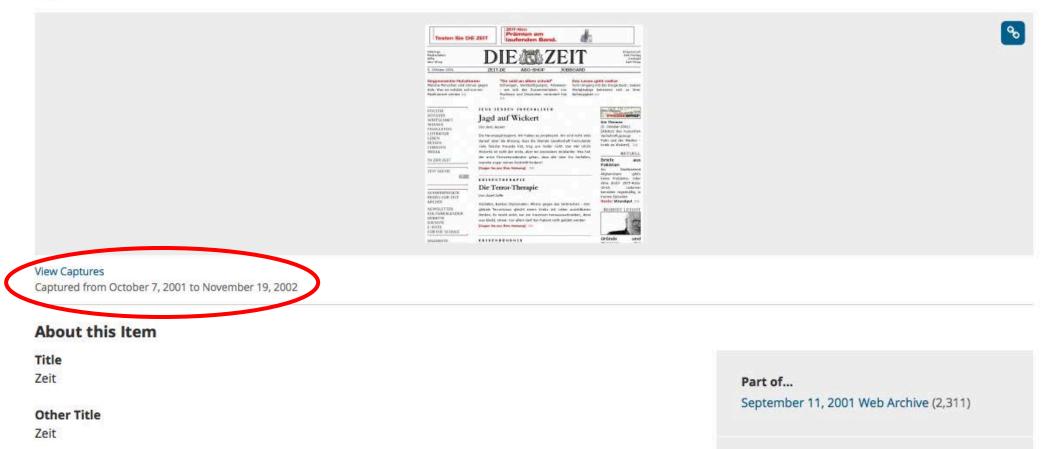


Subscribe Share/Save

#### ARCHIVED WEB SITE

Zeit

Zeit





15 16 17 18

29 30

20 21

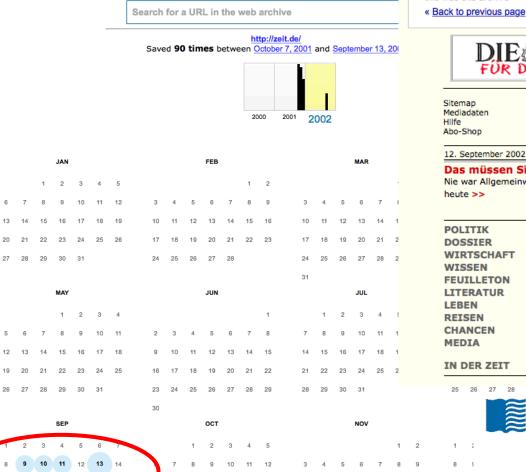
14 15 16 17 18 19

20 21 22 23 24 25 26

27 28 29 30 31

19

25





Note: External links, streaming audio/video, forms and search boxes may not function within this web site archive

15

29 3

25 26 27 28 29 30

18 19 20 21 22 23

24

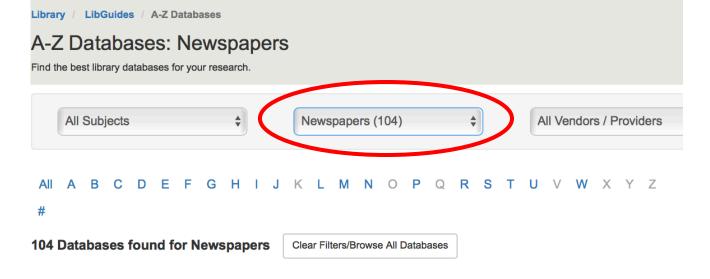
Library of Congress Web Archives Collection This is an archived Web site from the Library of Congress http://www.zeit.de/ Archived: 09/13/2002 at 04:22:52

#### ◄ First (10/08/2001) < Previous #89 of 90 Next ► Last (09/13/2002) ►</p>



### **Resource Not In Archive**

The Resource you requested is not in this archive. More options: Try Searching all pages under <u>http://www.zeit.de/</u> Ask a Librarian



#### Α

#### Access World News 🖸 DU

Provides extensive news coverage at any level—local, state, regional, national and international. Features the vast majority of U.S. newspapers by circulation, along with almost one thousand hard-to-find local and regional titles, and offers searchable news video clips and over two thousand international news sources from scores of countries on six continents, translated into English when written in other languages.

#### African American Newspapers, 1827-1998 C DU 🖧

Provides access to U.S. newspapers from more than 35 states chronicling a century and a half of the African American experience.

#### African Newspapers (World Newspaper Archive) C DU

Searchable African newspapers published in the 19th and early 20th centuries. Featuring English- and foreign-language titles from Angola, Ghana, Guinea-Bissau, Kenya, Lesotho, Liberia, Madagascar, Malawi, Mozambigue, Namibia, Nigeria, Sao Tome and Principe, Sierra Leone, South Africa, Swaziland, Tanzania, Uganda, Zambia and Zimbabwe.

#### America's Historical Newspapers C DU

Thousands of titles chronicle centuries of American history, culture and daily life. Includes titles from all 50 present states. This database provides access to digitized versions of primary source material.

#### American Business: Agricultural Newspapers C DU ぬ

Historical newspapers covering agriculture, agricultural technology, and economics of farming published between 1788 through 1894.

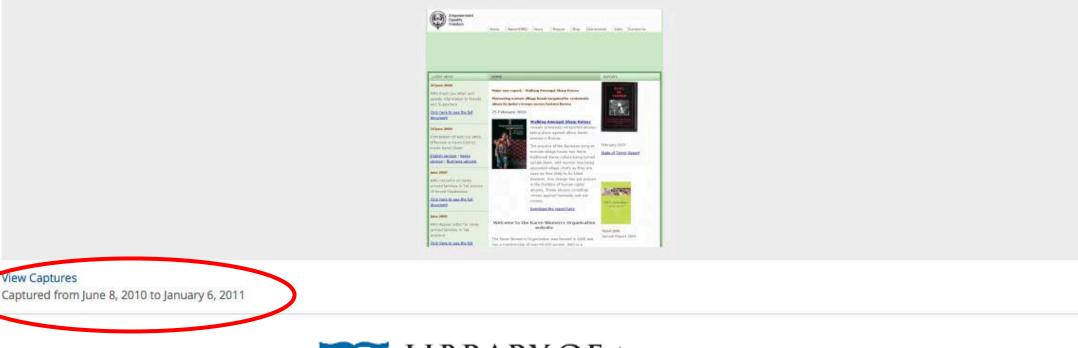
-

-4

4

-

# ARCHIVED WEB SITE Karen Women Organisation

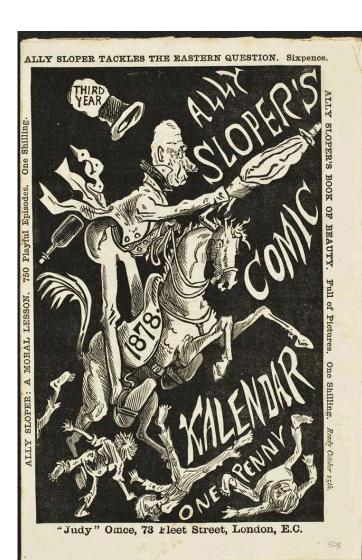




Search for a URL in the web archive

Q

Archived content not available outside of Library of Congress premises: this archived web site content displays only at the Library of Congress.



and apply Rirohu. That as proceeding by Martial Law can he had against Journal Streast, which would, and any Degree, meser the purposes of a Trial; Find as the Bears are in this Resolution, they are left to fellow the Time pointed out to theme by the first of Directory in the 53. Davagraph of this Letter by the Butnough Revolved thursfore, that General Awards' information from the service be continued, and that he be directed To force to bigland by the first proper convey ance. 3 Alow Coboact Jugin Cha! Oakeley during undo Oxtract Jeneral Orders 2. March 1778. The propension of Bigadien General Annal of the Tourient blowmil . Igut , Chat Oakeley Surg My the To the struck Themas Roundeld, Dig Proid ! H. B. ... V Jav orner. So? Currel of Fort S. George. : 1 attenthe Sir Sirs had the honor of securing your Letter under date the 27. " of last Month Legether with . Copy of the Terolution of the 24 had muth figured I. My Surstary Oakily and have five been fors and with be boards from Minutes of formalta tion of the 16" 23 & 24" flast Month Jones by of Agramment Jugned by me at the hidia House In the Letter you did me the honor of Adderstring our you are pleased to experit your due . , eititily on considering the particular distantion



How do libraries preserve the primary sources on the Internet?

- A constantly moving target
- Too much to ever preserve in depth
- Collaboration
  - To preserve some resources at the broad/surface level
    - Are the Internet Archive and International Internet Preservation Consortium doing this well enough?
  - To preserve some resources in depth
    - Is this the role for libraries?
- Legal and ethical issues

# Marijuana in Colorado

PotGuide.com

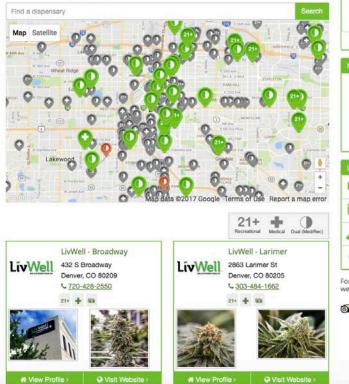
LOGIN | SIGN UP

STATES DISPENSARIES DEALS TRAVEL STRAIN PROFILES / REVIEWS MARIJUANA NEWS & CULTURE

#### Recreational and Medical Marijuana Stores in Denver

#### Home > Where to Buy Marijuana > Denver

Deriver is the largest city in Colorado, as well as the state capital. In Deriver you will find the highest concentration of recreational marijuana stores in Colorado, with a large number of select stores selling recreational and medical marijuana, 11 selling recreational only, and 71 medical only. Marijuana stores in Deriver are required to close by 7pm. You may want to explore the area on a <u>B-cycle</u>, a rental blke system offered throughout the city.







advertisement

Search Find a dispensary 1365 Map Satellite Derby 21+ wood Westminster W 76th ANE W 74th Ave X GroundSwell Groundswell Cannabis W 66th Ave Boutique 4th Ave 3121 E Colfax Ave Soth Ave Arvada Denver, CO 80206 Phone: 303-309-0078 NORTHEIELD V 44th Ave Website Store Profile Wheat Ride PARK HILL ©2017 Google Terms of Use Report a map error Groundswell Home Menu Current Specials Medical Recreational Products About GroundSwell Visit Us





#### Our St

Our business began in 2010 by serving those most in need of medicinal cannabis: homebound and invalid patients. Our center has evolved to meet the needs of all medical cannabis patients, and expanded in April 2015 to serve all adults over the age of 21.

The GroundSwell team is not new to the Bluebird District. In fact, we've played an integral part in the neighborhood's growth and improvements over the last 10 years. Our vision is to create a neighborhood resource that will educate and create awareness about responsible cannabis use and its medicinal benefits.



#### Our Mission

- To help every person who comes through our doors feel truly cared for, valued and respected,
- To protect cannabis patients by providing a safe and affordable alternative to the dangerous circumstances of the illegal drug market.
- To create an extraordinary environment of honesty and friendliness with a patient-centric approach.
- To improve and protect the safety, tranquility, and cleanliness of our immediate neighborhood.

## f University of Denver



# University of Denver 🔗

@uofdenver

## Home

About Posts Twitter

Videos

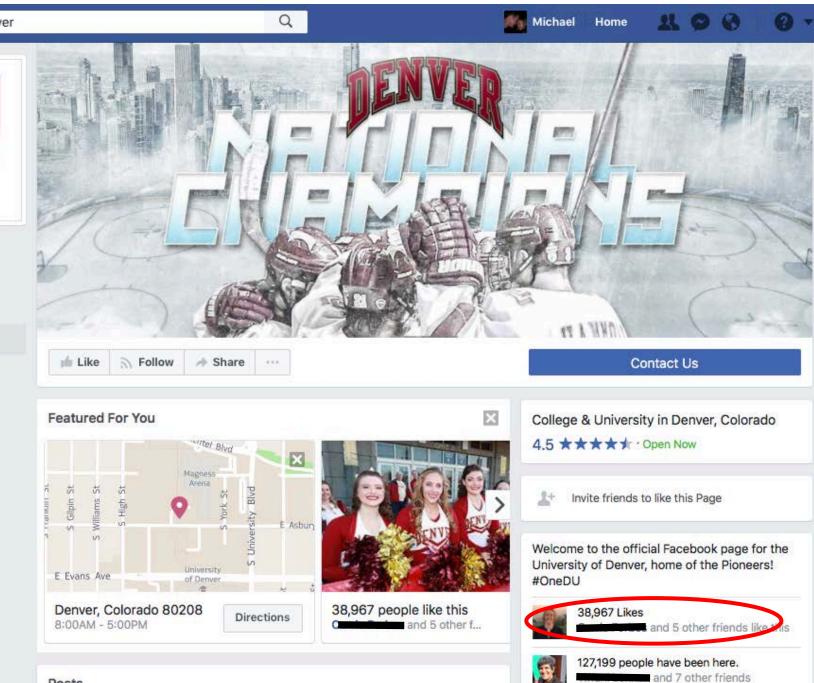
YouTube Photos

Events

Instagram

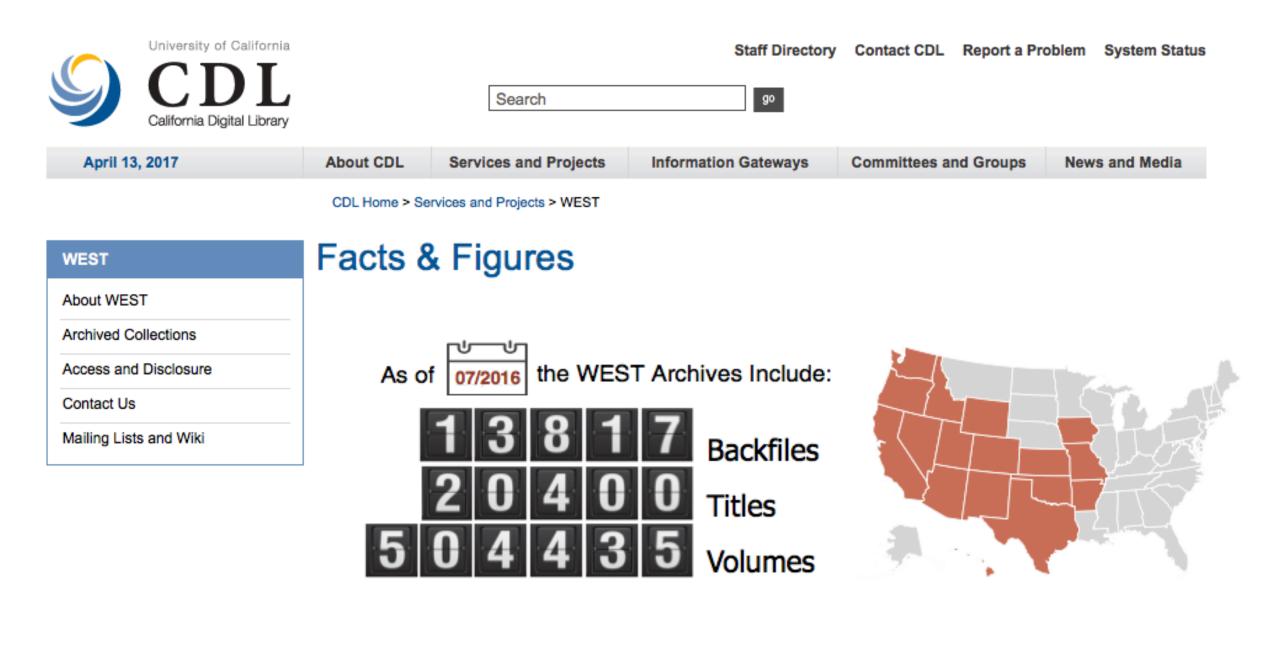
Reviews

Likes



Create a Page

Posts



## **OPEN ACCESS**

# Area-wide transition to open access is possible

A new study calculates a redeployment of funds in Open Access

April 27, 2015

The Max Planck Digital Library has put forward a study on the transformation of the subscription-driven system for scientific publications to an Open Access model. For the first time, quantitative parameters are presented showing that the liberation of scholarly literature is possible at no extra costs.

According to market analyses, annual turnovers of academic publishers amount to approximately EUR 7.6 billion. This money comes predominantly from publicly funded scientific libraries as they purchase subscriptions or licenses in order to provide access to scientific journals for their customers. Since more than a decade the Open Access movement, in which the Max Planck Society plays a major role, has been demanding free and immediate access to the results of academic research on the internet.

# To preserve the Internet, libraries should. . .

- Identify locally-important websites to preserve in depth
- Work collaboratively
- Expend a bit less money and effort on "traditional" resources