

Outline

- Web archiving basics
- Web archiving landscape
- Key approaches & challenges
- The Internet Archive
- Case studies
- Web archiving options for academic libraries
- Key considerations

Web archiving basics

- **What:** selecting, capturing, storing, preserving and managing access to Web resources over time
 - Notion of longevity, authenticity, reliability and future usefulness of archival material
 - Stewardship traditionally associated with libraries and archives
- **When:** started by the Internet Archive in 1996
- **How:**
 - Use crawler software to download Web resources automatically
 - Stored in archival formats
 - Provide access in Web archives by replaying archived content

Why and Who?

Why?

- Global information resource but ephemeral by nature
- Published national heritage: legal deposit
- Historical and sociological data that may not be found elsewhere: research value
- Government / corporate records
- Regulatory compliance, litigation readiness
- Preservation of scholarly citation

Who?

- Heritage and memory institutions
- Governments
- Scholarly organisations such as academic libraries
- Not-for-profit and commercial organisations
- Activists
- Individual researchers

HKUST Library homepage 1997 - 2016

The screenshot shows the Internet Archive Wayback Machine interface for the URL <http://library.ust.hk>. The page indicates that the site was saved 631 times between December 10, 1997, and May 2, 2016. A calendar view highlights the date May 2, 2016, as the most recent save. A bar chart at the top shows the frequency of saves over time, with a significant peak in 2004 and a smaller peak in 2016.

INTERNET ARCHIVE
WayBackMachine

<http://library.ust.hk>
Saved **631 times** between [December 10, 1997](#) and [May 2, 2016](#).

PLEASE DONATE TODAY. Your generosity preserves knowledge for future generations. Thank you.

1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 **2016**

Month	Day	Saves			
JAN	1	0			
	2	0			
	3	0			
	FEB	1	0		
		2	0		
		3	0		
		MAR	1	0	
			2	0	
			3	0	
			4	4	
			5	0	
			APR	1	0
				2	0
				3	0
				4	0
				5	0
				6	0
				7	0
				8	0
				9	0
				10	0
				11	0
				12	0
				13	13
				14	14
				15	0
				16	0
				17	0
				18	0
				19	0
				20	0
21				0	
22				0	
23				0	
24	0				
25	0				
26	0				
27	0				
28	0				
29	0				
30	0				
31	0				
MAY	1	2			
	2	0			
	3	0			
	4	0			
	5	0			
	6	0			
	7	0			
	8	0			
	9	0			
	10	0			
	11	0			
	12	0			
	13	0			
	14	0			
	15	0			
	16	0			
	17	0			
	18	0			
	19	0			
	20	0			
	21	0			
	22	0			
	23	0			
	24	0			
	25	0			
	26	0			
	27	0			
	28	0			
	29	0			
	30	0			
	31	0			
JUN	1	0			
	2	0			
	3	0			
	4	0			
	5	0			
	6	0			
	7	0			
	8	0			
	9	0			
	10	0			
	11	0			
	12	0			
	13	0			
	14	0			
	15	0			
	16	0			
	17	0			
	18	0			
	19	0			
	20	0			
	21	0			
	22	0			
	23	0			
	24	0			
	25	0			
	26	0			
	27	0			
	28	0			
	29	0			
	30	0			
	31	0			
JUL	1	0			
	2	0			
	3	0			
	4	0			
	5	0			
	6	0			
	7	0			
	8	0			
	9	0			
	10	0			
	11	0			
	12	0			
	13	0			
	14	0			
	15	0			
	16	0			
	17	0			
	18	0			
	19	0			
	20	0			
	21	0			
	22	0			
	23	0			
	24	0			
	25	0			
	26	0			
	27	0			
	28	0			
	29	0			
	30	0			
	31	0			
AUG	1	0			
	2	0			
	3	0			
	4	0			
	5	0			
	6	0			
	7	0			
	8	0			
	9	0			
	10	0			
	11	0			
	12	0			
	13	0			
	14	0			
	15	0			
	16	0			
	17	0			
	18	0			
	19	0			
	20	0			
	21	0			
	22	0			
	23	0			
	24	0			
	25	0			
	26	0			
	27	0			
	28	0			
	29	0			
	30	0			
	31	0			
SEP	1	0			
	2	0			
	3	0			
	4	0			
	5	0			
	6	0			
	7	0			
	8	0			
	9	0			
	10	0			
	11	0			
	12	0			
	13	0			
	14	0			
	15	0			
	16	0			
	17	0			
	18	0			
	19	0			
	20	0			
	21	0			
	22	0			
	23	0			
	24	0			
	25	0			
	26	0			
	27	0			
	28	0			
	29	0			
	30	0			
	31	0			
OCT	1	0			
	2	0			
	3	0			
	4	0			
	5	0			
	6	0			
	7	0			
	8	0			
	9	0			
	10	0			
	11	0			
	12	0			
	13	0			
	14	0			
	15	0			
	16	0			
	17	0			
	18	0			
	19	0			
	20	0			
	21	0			
	22	0			
	23	0			
	24	0			
	25	0			
	26	0			
	27	0			
	28	0			
	29	0			
	30	0			
	31	0			
NOV	1	0			
	2	0			
	3	0			
	4	0			
	5	0			
	6	0			
	7	0			
	8	0			
	9	0			
	10	0			
	11	0			
	12	0			
	13	0			
	14	0			
	15	0			
	16	0			
	17	0			
	18	0			
	19	0			
	20	0			
	21	0			
	22	0			
	23	0			
	24	0			
	25	0			
	26	0			
	27	0			
	28	0			
	29	0			
	30	0			
	31	0			
DEC	1	0			
	2	0			
	3	0			
	4	0			
	5	0			
	6	0			
	7	0			
	8	0			
	9	0			
	10	0			
	11	0			
	12	0			
	13	0			
	14	0			
	15	0			
	16	0			
	17	0			
	18	0			
	19	0			
	20	0			
	21	0			
	22	0			
	23	0			
	24	0			
	25	0			
	26	0			
	27	0			
	28	0			
	29	0			
	30	0			
	31	0			

HKUST Library homepage 10 Dec. 1997

INTERNET ARCHIVE
WayBackMachine

http://library.ust.hk/

NOV **DEC** JAN
1996 **10** 1999

631 captures
10 Dec 97 - 2 May 16


[Close X](#)

[Help ?](#)



Welcome to the HKUST Library's web server. Please send us your comments and suggestions. Write to us at library@usthk.ust.hk If you are looking for information about the University, see the [HKUST Home Page](#).

- Featured Site: [Engineering Information Village](#) | [IBM Patent Server](#)
- Featured UST Library pages: [Gems of Contemporary Chinese Prints](#)
- Databases Under Evaluation: [Congressional Compass](#) | [MathSciNet](#)
- HKUST Library [News and Events](#) | [Seminar on Chinese Information Processing](#)

 [About the Library](#)

 [Library Resources](#)

 [Frequently Asked Questions](#)

 [Search This Server](#)

 [Library Services and Programs](#)

 [HKUST Manuals and Databases](#)

 [Beyond the Library](#)

 [Statistics for This Server](#)

rev. 28 November 1997

The HKUST Library cannot guarantee the accuracy or availability of any linked resources.

HKUST Library homepage 28 May 2016

The screenshot shows the HKUST Library homepage in a web browser window. The browser's address bar displays the URL: `web.archive.org/web/20160528233941/http://library.ust.hk/`. The browser's tabs include "HongKong - Google Slides", "Academic Librarian 4", "#AL42016 hashtag on Twitter", and "HKUST Library Home Page". The browser's address bar also shows "633 captures" and "10 Dec 97 - 28 May 16". The browser's status bar shows the date "28" and the year "2016".

The main content area features a green header with the library's name in Chinese: "香港科技大學圖書館" and in English: "The Hong Kong University of Science and Technology Library". The header also includes social media icons for Facebook, Weibo, YouTube, and RSS. Below the header, there are navigation links: "Hours", "Forms", "Ask A Librarian", "My Account", "HKALL", "Interlibrary Loan", "Reserve Rooms/PCs", "Learning Commons", and "Contact Us".

The main content area is divided into several sections:

- Mobile Version**: A link to the mobile version of the site.
- Featured Resources**: A list of featured resources including "Met Opera On Demand", "Gale Virtual Reference Library", and "民國時期期刊全文數據庫".
- Instruction**: A list of instructional resources including "CENG 1600", "LANG 2030", "LANG 2070", and "Event & Registration".
- Opinions**: A list of opinion resources including "Trial E-Resources: SCMP ePaper" and "Library Survey Results".
- Events**: A list of events including "Map Exhibition: 'Tartary' from Marco Polo to the Enlightenment", "Academic Librarian 4 Conference", "Job Vacancies for Student Helpers", and "HKUST 25th Anniversary Exhibition".
- Projects**: A list of projects including "Library Readers ALERT: Hyperloop is Coming", "Collection Highlights - Award Winning Books", "China in European Maps: Catalog", and "Library Outstanding Student Helpers".

The main content area also features a search bar with the text "Search the Library Catalog & Primo Central (About)". The search bar includes a "Word/Phrase" dropdown, a "Quick Search" input field, and checkboxes for "Library Catalog" and "Primo Central". The search bar also includes a "Search" button and links to "Go to the Classic Catalog", "Collection Highlights", "New Acquisitions", and "My Account".

The main content area also features a "PowerSearch" section with tabs for "Databases", "E-Journals", "Reserves", and "Guides". Below the search bar, there is a "Collections & Resources" section with a list of resources including "Library Collections", "Research Guides", "Finding:", "Books", "Articles & Journals", "Theses", "Audio & Videos", and "More:". The "More:" section includes "DataSpace@HKUST", "E-Books", "E-Learning Videos & Games", "E-Theses", "HKUST Digital Images", "Institutional Repositories (HK)", "Institutional Repository (HKUST)", "Internet Resources", "News Clippings", "Publications and Newsletters", "Rare & Special e-Zone", "Scholarly Publications", "Special Collections", "Streaming Videos", and "University Archives".

The main content area also features a "Services & Information" section with a list of services including "Services for:", "Faculty & Staff", "Postgraduates", "Undergraduates", "Alumni", "Visitors & Others", "Services A-Z", "About Us", "News Blog", and "HKUST Publications". Below the "Services & Information" section, there is a "Frequently Asked Questions" dropdown and a "Search library.ust.hk" input field with a "Site Search" button.

The main content area also features a "Check IC Availability" section with a graphic showing a grid of red and green squares representing the availability of various resources.

The main content area also features a "New tall furniture along the windows on 1/F" section with a photo of a library interior and a "15 / 39" indicator.

The footer of the page includes a "WhatsApp : 9701-1055" link, a "Sitemap Feedback Statistics" link, and social media icons for Facebook, Weibo, YouTube, and RSS.

Web archiving landscape (1)

- Internet Archive's Wayback Machine
 - Largest and oldest Web archive in existence
 - Global scale, unrestricted access
- National Web archives
 - Operated by national libraries & organisations
 - Goal is to preserve national heritage
 - 30+ national libraries have web archiving programmes
- Government and corporate Web archives
 - Preserving records
 - [UK Government Web Archive](#)
 - [Websites Archives of EU Institutions](#)

Web archiving landscape (2)

- Research-driven Web archives
 - Mostly operated by university libraries
 - [The Human Rights Web Archive](#)
 - [Digital Archive for Chinese Studies](#)
- Efforts of individuals and small groups
 - Archive Team
 - Dale Askey Archive
- Web archiving service providers
 - [Archive-It](#) & [Internet Memory Foundation](#)
 - Collections open to everyone
- Commercial services: e.g. [Hanzo](#)
 - Closed
- International Internet Preservation Consortium ([IIPC](#))
 - High concentration of use of technology and practice

Collecting strategy

- Broad crawls of global or national Web
 - Cannot collect everything
 - Arbitrary boundaries
- Selective archiving
 - Subset of a national domain, e.g. .gov.uk
 - Thematic, events-based collections
 - Focuses on resources considered to be particularly valuable or relevant to an organisation's collecting remit
 - Requires additional resource
- Many use a combined strategy of domain and selective archiving

Open source technology

- **Heritrix:** Web crawler – crawls and captures web pages
- **Umбра:** Assists the crawler to access social media and other sites in the same way a browser would
- **Wayback:** Access tool for rendering and viewing pages - surf the web as it was
- **NutchWAX & SOLR:** Full-text and metadata search engine
- **WARC:** ISO standard for storing Web archives

Who archived what? - Memento look-up service

UK Web Archive Home Mementos

http://www.ust.hk Find Mementos...







URL: <http://www.ust.hk>

Snapshots: 1274 in 6 archive(s)

Date Range: 1996 to 2016 (20 days ago)

Request Archive: via the UK Web Archive, via WebCite™

Host Chart Host Table

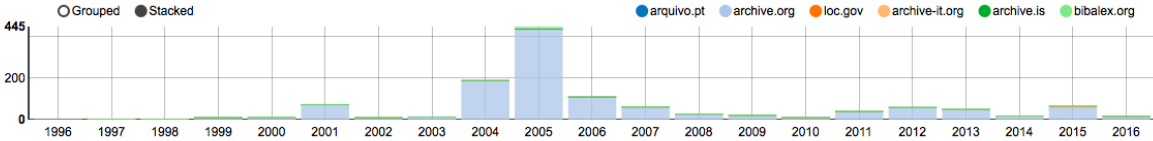
Archive	Snapshots
 archive.org	1245
 bibalex.org	15
 archive.is	9
 loc.gov	2
 archive-it.org	2
 arquivo.pt	1

1996 2005 2016 LIVE

Snapshot Chart Snapshot Table

○ Grouped ● Stacked

● arquivo.pt ● archive.org ● loc.gov ● archive-it.org ● archive.is ● bibalex.org



Bookmarklet: [Find Mementos](#).

For more information, see [the Mementos homepage](#).

This web interface uses the Memento aggregate TimeGate hosted by [lanl.gov](#).

For more information on Memento, see www.mementoweb.org.

Conceptual challenges

- Existing concepts, legislation, network...
- Different from traditional archives, not just because it contains digital content
- Different from “live Web” and digitised material, considered as “reborn” by [Brugger & Finnemann](#)
- Blurred distinction between “publications”, “grey literature”, and “records” - confusion around responsibilities
- Additional boundary issues
 - The Web never ends
 - National approach to archiving global Web

Technical challenges

- Content gaps in Web archives
 - JavaScript, streaming & downloadable media, password protected & paywall content, form and database driven content, robots.txt exclusions
 - Social media: Twitter, YouTube, Facebook...
- Scale
 - Full-text search
- Complexity throughout the end to end web archiving process
 - Temporary inconsistency
- Benchmarking
 - How do you know if you have done a good job?
- In an arms race with Web technology

Legal challenges

- Copyright
- Libellous or defamatory content
- Illegal content
- Personal data
- “Rights to be forgotten”
- Robots.txt
- Notice and take-down

Curatorial & organisational challenges

- Web archiving requires a hybrid of skills
 - Not supported by existing organisational structure
- Collaboration adds complexity
- Selection
 - Bias?
 - Gaps and overlap
- Authenticity
- Curation largely manual & does not scale up
 - Quality Assurance (QA)
 - Descriptive metadata
- Growing costs

Access & use challenges

- Comprehensiveness versus openness
- URL-based discovery, focus on replaying “text”
- Very little is explained or explicitly stated beyond HTML
 - [The Unknown Aspects of Web Archives](#)
- Not enough versus too much
 - Pre-selected or defined collections not relevant to all researchers; difficulty in finding relevant content in large scale web archive
- Methodological challenges
 - [Revealing British Euroscepticism in the UK Web Domain and Archive Case Study](#)
 - Using web archive as primary source: speculative and experimental
 - Unexpected pitfalls and challenges
 - Digital methodology: necessary requisite?
 - Google mindset



The Internet Archive:

Non-Profit Library

Universal Access to All Knowledge

90,000 Software Titles

2,000,000 Moving Images

2,300,000 Book Archive

2,400,000 Audio Recordings

3,000,000 Hours of

4,000,000 Television

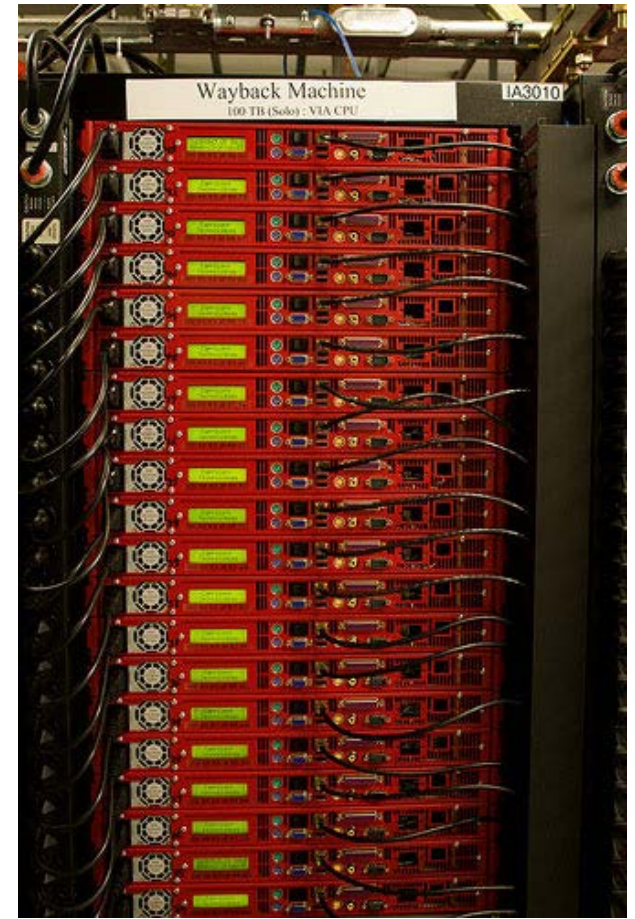
472,000,000,000 eBooks

Web Captures

25,000,000,000,000,000 Bytes Archived

(25 PetaBytes)

Petabox storage system



Web scale preservation

- Ongoing crawls incl. snapshot of the global web; wikipedia, news, RSS feeds, YouTube etc.
 - 1 billion URLs/week

- Save Page Now

- Also receives donation of crawl data from comp



- Wayback Machine, access service to Internet Archive's Web archive collections

- Released in 2001
- 480+ Billion URLs, 100+ million websites

- Content in 40+ Languages

- 600,000 visitors / day

- APIs https://archive.org/help/wayback_api.php



Libraries and archives partnership

- Have worked with many national libraries / archives since 2000
- Crawling of national domains and building thematic collections
- Tailored to partners' requirements
- Identify and collect in-scope content outside country code top level domains (ccTLDs)
- Comprehensive coverage of national domain, e.g 546 million unique .nz URLs for NLNZ versus 273 million through other crawls
- Deliver bulk data + metadata, index, reports and datasets



Support Open Source Software

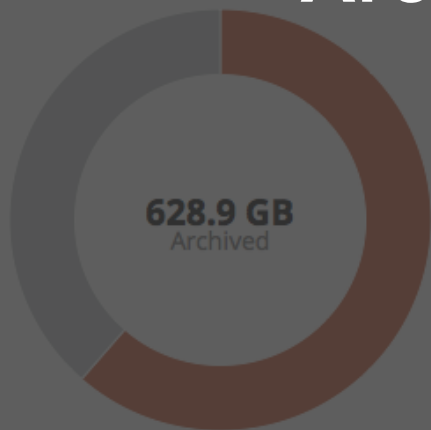
HERITRIX



Archive-It Demo Account

3.6 TB archived since Sep 22, 2010

Current Subscription



Data Budget Usage

Active Collection List (3 Active Collections)

[Create a Collection](#)[Download Collection List](#)

Collection Name	Data (this period)	Docs (this period)	Active Seeds	Last Crawl
Archive-It Websites and Press	72.4 GB	1,187,978	57	Jan 6, 2016
2015 United Nations Climate Ch	42.6 GB	726,589	5	Dec 10, 2015
Climate Change	1.9 GB	92,932	13	Dec 27, 2015

Archive-It: ait@archive.org

Archive-It

- Subscription service launched in February 2006
- Fully hosted web application for creating, managing accessing and storing web archive collections
- Captures HTML, PDF, Twitter, YouTube...
- Support for cataloguing and metadata
- Browse archived content 24 hours after capture; full-text search available within 7 days
- Private access option



Archive-It partners

- 400+ partners worldwide
 - Including universities and colleges, state archives and libraries, museums and art galleries, public libraries and local governments, national organisations, NGOs...
- **Motivation**
 - Create a thematic/topical web archive on a specific subject or event
 - Often related to traditional collecting activity around the same focus
 - Fulfill a mandate to capture/preserve evolving web history
 - Meet records retention requirements
 - Capture publications/documents that are no longer in print form
 - Construct a historical record of an institution or individual's web/social media presence over time



University of Texas at Austin: LAGDA

- Archive government documents from 18 different countries in Latin America
- Content includes:
 - Full-text versions of official documents
 - Original video and audio recordings of key regional leaders
 - Thousands of annual and "state of the nation" reports
 - Specific collections for Latin American elections and political parties

Honduras Presidential site 2008 (before the Coup)

You are viewing an archived web page, collected at the request of [University of Texas at Austin Libraries, Latin American Government Documents Archive](#) using [Archive-It](#). This page was captured on 22:37:09 Dec 15, 2008, and is part of the [Latin American Government Documents Archive, LAGDA](#) collection. The information on this web page may be out of date. See [All versions](#) of this archived page. hide

Casa Presidencial HONDURAS

Luchemos y salvemos la naturaleza, el momento es hoy, mañana será demasiado tarde



[Pagina Principal](#) [Correo Institucional](#)

Hoy es :
December 15, 2008
[Menu Principal](#)
Navegación
[Presidente Zelaya](#) >>
[Gobierno](#) >>
[Multimedia](#) >>
[Prensa](#) >>
[En Contacto](#) >>
[Para Tu Informacion](#) >>
[Presidente Zelaya](#)
[HONDURAS: PAIS ABIERTO A LA INVERSION EXTRANJERA](#)
[HONDURAS: A COUNTRY OPEN TO FOREIGN INVESTMENT](#)

YouTube Videos Logros Poder Ciudadano

Petrocaribe y Petroalimentos

[Principal](#) [Encuesta](#) [Video de la Semana](#) [Logros](#)

ULTIMAS NOTICIAS



En 183 Aniversario del Ejército PRESIDENTE ZELAYA DESTACA NUEVO ROL DE LAS FUERZAS ARMADAS
Tegucigalpa. En el marco del 183 aniversario del Ejército de Honduras, el Ministro de Defensa, Aristides Mejia, anunció una inversión de 20 millones de lempiras para restaurar la sede de los batallones, la ampliación del hospital militar del Norte y la presentación de un nuevo uniforme para los hombres de verde olivo.
NEW

CANAL 8 - RED INFORMATIVA



DXV Asamblea en New York (ONECA)

Usted es el Visitante Numero :
157,456

EJECUTIVO RESPALDA CUALQUIER ACTO DE RECUPERACION DE LAS

Honduras Presidential site in 2009 (during the Coup)

You are viewing an archived web page, collected at the request of University of Texas at Austin Libraries, Latin American Government Documents Archive using [Archive-It](#). This page was captured on 20:32:26 Sep 19, 2009, and is part of the [Latin American Government Documents Archive, LAGDA](#) collection. The information on this web page may be out of date. See [All versions](#) of this archived page.

hide

Red Hat Enterprise Linux Test Page

This page is used to test the proper operation of the Apache HTTP server after it has been installed. If you can read this page, it means that the Apache HTTP server installed at this site is working properly.

If you are a member of the general public:

The fact that you are seeing this page indicates that the website you just visited is either experiencing problems, or is undergoing routine maintenance.

If you would like to let the administrators of this website know that you've seen this page instead of the page you expected, you should send them e-mail. In general, mail sent to the name "webmaster" and directed to the website's domain should reach the appropriate person.

For example, if you experienced problems while visiting [www.example.com](#), you should send e-mail to "webmaster@example.com".

For information on Red Hat Enterprise Linux, please visit the [Red Hat, Inc. website](#). The documentation for Red Hat Enterprise Linux is [available on the Red Hat, Inc. website](#).

If you are the website administrator:

You may now add content to the directory `/var/www/html/`. Note that until you do so, people visiting your website will see this page, and not your content. To prevent this page from ever being used, follow the instructions in the file `/etc/httpd/conf.d/welcome.conf`.

You are free to use the image below on web sites powered by the Apache HTTP Server:

[[Powered by Apache](#)]

Honduras Presidential site in 2010 (after the Coup)

Presidente Lobo | Rosa Elena de Lobo | Noticias | Enlaces | - x +

 Portal de Transparencia

[Inicio](#) [Presidente de Honduras](#) [Biografía del Presidente Lobo](#) [Primera Dama](#) [Designados](#) [Bono 10mil](#) [Contactenos](#)

Select Language ▾



Sociedad civil acuerpa esfuerzos del Gobierno por lograr acuerdo con el FMI
Agenda del Ejecutivo
***Ei...
◀ 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 ▶ Play


COMUNICADO

COMUNICADO
La Secretaría de Relaciones Exteriores, a la opinión pública nacional e internacional hace saber que el gobierno de Honduras ejecuta un programa de cooperación antidrogas con el gobierno de los Estados Unidos de América.

- DILMA ROUSSEFF: "EN BRASIL LA SUSTENTABILIDAD NO QUEDARÁ A MERCED DE LAS CRISIS"
- Presidente Lobo Sosa agasaja a niños del proyecto "Esperanza para todos"
- Presidente Lobo Sosa juramenta a joven embajador en el programa Ruta Quetzal
- Presidente Lobo Sosa nombra al nuevo director adjunto...
- Presidente Lobo Sosa recibe reconocimiento como ciudadano distinguido de Trujillo
- Presidente Lobo Sosa continúa entrega del Bono 10 Mil en Olancho

Search 

AGENDA PRESIDENTE PORFIRIO LOBO

There are no events at this time

SECRETARÍAS DE ESTADO

 SECRETARÍA DE CULTURA, ARTES Y DEPORTES Honduras	 FONDO NACIONAL DE GOBIERNECENCIA
 SECRETARÍA DEL INTERIOR Y POBLACIÓN Honduras	 SECRETARÍA DE EDUCACIÓN Honduras
 SECRETARÍA DE DEFENSA NACIONAL Honduras	 SECRETARÍA DE AGRICULTURA Y GANADERÍA Honduras

Columbia University Libraries: [the Human Rights Web Archive](#)

- Based at the Center for Human Rights Documentation & Research, Columbia University Libraries
- Existing commitment to collecting physical archival material from human rights organisations and activists
- Also wanted to collect born-digital outputs
- At-risk status of human rights organisations and their websites
- Received support from Mellon Foundation: exploratory, implementation and project grants
- Mainly funded by the University since 2013
- Sustainability
 - Mellon support allowed for dedicated staff
 - Existing expertise and engagement
 - Commitment to maintaining and growing collection

HUMAN RIGHTS WEB ARCHIVE

CENTER FOR HUMAN RIGHTS DOCUMENTATION & RESEARCH AT COLUMBIA UNIVERSITY

BETA

[Home](#)
[LEARN MORE](#)
[NOMINATE A SITE](#)
[CONTACT](#)

SEARCH HISTORY

Showing item 1 of 1.

Amnesty International



Creator	Amnesty International. International Secretariat
Organization Type	Non-governmental organizations
Organization Based In	United Kingdom
Geographic Focus	Global focus
Subjects	Human rights ; Capital punishment ; Torture ; Civil rights ; Political persecution ; Political prisoners
Summary	Provides information on various aspects of human rights issues: latest news and information by country and region, photos, and, AI campaigns. Includes information about the organization: mission statement, annual reports, urgent actions, and library (archived reports, news releases). Access provided to a search engine, and links to related sites.
Languages	English ; Spanish ; French ; Arabic



[View site captures for http://www.amnesty.org](#)

[Email This](#)

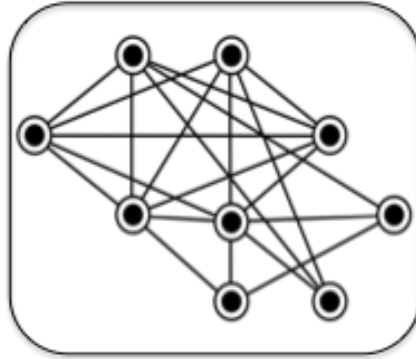
[Contact](#) • [Problem Report](#) • [Terms of Use](#)

© Columbia University Libraries 2012

Web Archive Datasets



WAT Datasets
(Web Archive
Transformation)
Key Metadata from Every
Resource



LGA Datasets
(Longitudinal
Graph Analysis)
What Links to What
over Time



WANE Datasets
(Web Archive
Named Entities)
Names of People, Places,
Organizations

October 2014 entities – Columbia U.'s Human Rights collection (~4M docs; 330GB)

TOP 6 People Entities (of 1,887,523)



Barack Obama



Ilham Tohti



Julien Assange



George W. Bush



Pu Zhiqiang



Bashar al-
Assad




TOP 6 Organisations (of 1,335,232)



TOP 9 Locations (of 373,155)





ArchiveWeb Interface


ArchiveWeb  [My resources](#) [Group overview](#) [Profile](#)  


back to [Human Rights D...](#) [Web](#) [Images](#) [Videos](#)


Service ▾ Date ▾ Group ▾ Collector ▾ Author ▾ Coverage ▾ Publisher ▾ Tags ▾


International Criminal Tribunal for the former Yugoslavia
ArchiveWeb » <http://www.icty.org/>
Captured between 3/22/12 and 7/2/15
The International Criminal Tribunal for the former **Yugoslavia** (ICTY) was established by Security Council resolution 827. This resolution was passed on 25 May 1993 in the face of the serious
Options:  [Copy to My resources](#) **Group:** [Human Rights](#); **Collector:** Columbia University Libraries, Center for Human Rights Documentation and Research

Youth Initiative for Human Rights
ArchiveWeb » <http://yhr.org/>
Captured between 4/29/09 and 10/27/15
enhance the participation of youth in the processes of transitional justice by promoting the truth about the wars in the former **Yugoslavia**. Site contains reports and newsletters for download, news, press
Options:  [Copy to My resources](#) **Group:** [Human Rights Documentation Initiative](#); **Collector:** UT Libraries


Yugoslavia - Wikipedia, the free encyclopedia
<https://en.wikipedia.org/wiki/Yugoslavia>
Yugoslavia (Serbo-Croatian, Macedonian, Slovene: Jugoslavija, Југославија) was a country in Southeast Europe during most of the 20th century.
Options:  [Copy to My resources](#)

The Breakup of Yugoslavia, 1990–1992 - State
<https://history.state.gov/milestones/1989-1992/breakup-yugoslavia>
The Breakup of **Yugoslavia**, 1990–1992. Issued on October 18, 1990, National Intelligence Estimate (NIE) 15–90 presented a dire warning to the U.S ...
Options:  [Copy to My resources](#)

History of the Former Country of Yugoslavia
<http://geography.about.com/od/politicalgeography/a/fmyugoslavia.htm>
Learn about the former European country of **Yugoslavia** (1945-1992), now composed of Slovenia, Macedonia, Croatia, Serbia and Montenegro, and Bosnia, from your expert ...
Options:  [Copy to My resources](#)

Yugoslavia - Britannica.com
<http://www.britannica.com/place/Yugoslavia-former-federated-nation-1929-2003>
Yugoslavia, Yugoslavia: historical boundaries, 1919–1992 Encyclopædia Britannica, Inc. former federated country situated on the west-central Balkan Peninsula.
Options:  [Copy to My resources](#)

Yugoslavia - New World Encyclopedia
<http://www.newworldencyclopedia.org/entry/Yugoslavia>
Yugoslavia describes three political entities that existed one at a time on the Balkan Peninsula in Europe, during most of the twentieth century.



Belgrade

Belgrade is the capital and largest city of Serbia. The city is located at the confluence of the Sava and Danube rivers, where the Pannonian Plain meets the Balkans. Its name in English translates to White city. The city proper has a population of over 1.1 million, while its metropolitan area has over 1.6 million people, making it one of the largest cities in Southeast Europe.

Country : Serbia
Population (as of 2011) : 1232731
Time Zone : Central European Time

Related Links

- [Yugoslavia](#)
- [Socialist Federal Republic of Yugoslavia](#)
- [Kingdom of Yugoslavia](#)
- [Serbia and Montenegro](#)

Group overview

Resources of Human Rights Documentation Initiative (269 Resources)

Join group

Overview Resources Members Presentations Links Forum

Date ▾

Human rights - Wikipedia, th...



Text added by iadem04

Human rights - Wikipedia, th...



Text added by iadem01

Radio Miraya | Facebook



Text added by Admin

Bekhsoos | Facebook



Text added by Admin

SAACID



Text added by Admin

Buraku kaiho domei zenkok...



Text added by Admin

Fond "V zashchitu prav zakl...



For Human Rights, All-Russi...



Youth Initiative for Human RI...



Text added by Admin

Youth Initiative for Human Rights


<http://yihr.org/>

Author: Youth Initiative for Human Rights (YIHR)

Source: Archive-It

Coverage: Europe--Bosnia and Herzegovina; Europe--Serbia; Europe--Montenegro; Europe--Kosovo, Eurasia

Publisher: Youth Initiative for Human Rights (YIHR)

Collector: UT Libraries

Group: Human Rights Documentation Initiative

Date: July 21, 2015

Description

In English, Serbo-Croatian and Albanian. Formed in 2003, YIHR aims to provide protection for the victims of violations of human rights, establish new connections between the post-war generations and enhance the participation of youth in the processes of transitional justice by promoting the truth about the wars in the former Yugoslavia. Site contains reports and newsletters for download, news, press releases, information on programs, and individual pages for YIHR offices.

View archived versions (36): [List view](#) [Timeline view](#)

Tags

Restoring a lost domain

By Dr Anat Ben-David (@anatbd) of the [Open University of Israel](#)

Archival reconstruction of the Yugoslav Internet domain

.yu deleted from the internet root zone on 20 March 2010

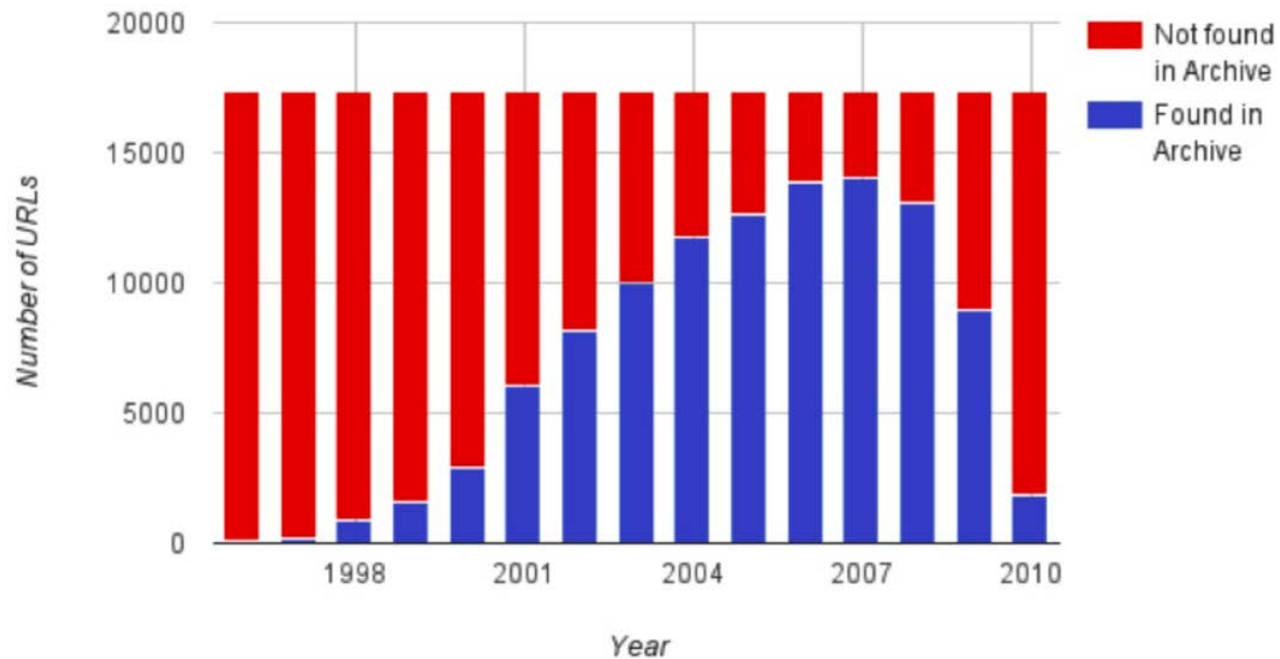
Hosted 70,000 websites at its peak

Used various sources to generate a list of URLs in .yu domain incl.
wikipedia

Extracted outlinks to .yu from archived pages in the Wayback
Machine

Recovered dataset contains 1.92 million unique pages once hosted
in the .yu domain between 1996 and 2010

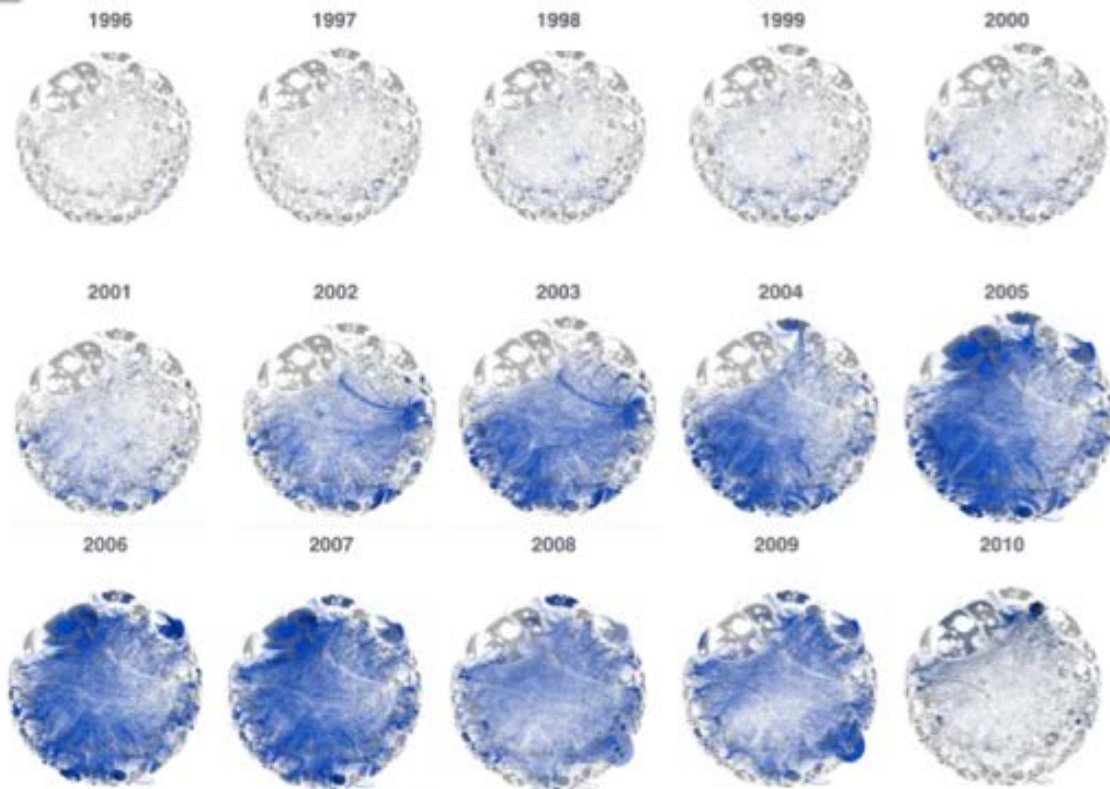
Archival coverage of the reconstructed dataset



.yu
Other



The networked structure of the .yu domain, 1996-2010



More information

- On reconstruction of .yu

<https://webarchivehistorians.org/2015/09/07/what-does-the-web-remember-of-its-deleted-past/>

<http://www.slideshare.net/anatbd/resaw2015-bendavid-presentation>

<http://www.bbc.co.uk/programmes/p031zp5y> 00:34 - 7:09

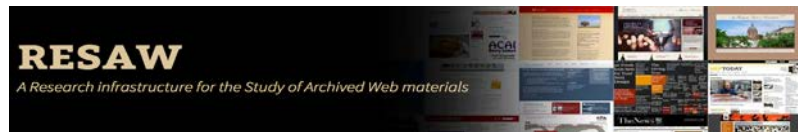
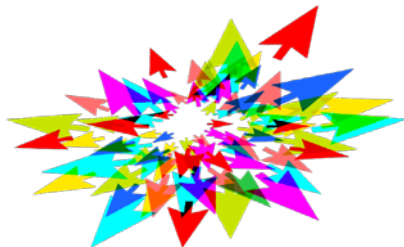
- [Web Archives for historians](#)

Historians who use, think about and work with web archives

Typology of research interest

- **Documentary:** Evidentiary, Attestation, Legal discovery/claim
- **Social/Political Scientists:** Communications, Politics/Government, Social Anthropology
- **Web Science:** Technology Systems and Protocols
- **(Digital) Humanities:** Historians and humanities disciplines, networks, collection building
- **Computer Science:** Information Retrieval, Data Processing and Indexing, Infrastructure and tools
- **Data Analysts:** Mining/Training, language processing, trend analysis, NLP/NER

Sample Research Partnerships



#iipcwac16

The academic library has died?

- The academic library has died," ... "One reason for cause of death is that library buildings were converted into computer labs, study spaces and headquarters for informational-technology departments." - Brian Sullivan, librarian at Alfred University
- No denying the academic library as we know is changing
- Still need to support teaching, learning and research
- Cannot walk away from the Web

Web archiving options

- In-house operation
 - High entry barrier
 - Maintenance of technical infrastructure
 - Ongoing tools development & user support
 - Staff expertise and training
- Outsource
 - Few service providers
 - Less control
- In-house + outsource
 - Operate core components in-house
 - Use existing web archives (e.g. API)
- Collaboration
 - E.g. centralised infrastructure and technical services, distributed collection development

Key considerations

- Make informed decision on why / why not?
- Do your homework: e.g.
 - [Web Archiving Environmental Scan](#)
 - DPC Technology Watch Report: [Web Archiving](#)
- What resources are available?
 - Curatorial and technical
- Determine a collecting strategy
 - Examine existing collection policies
- Determine an operational model
 - Outsource + collaboration allows you to gain knowledge and experience - part of learning process
- Appraise tools and services
- Know your rights and risks
- Run pilot, start small and progress
- Consider access, use, long term preservation and SUSTAINABILITY

Web archiving webinar series



LATEST ARTICLES

Wednesday, 18 May 2016

Four Recordings Cover Current Aspects of Web Archiving

[Learn More »](#)

Wednesday, 18 May 2016

Key LGBT Material Now Online Through Culture in Transit

[Learn More »](#)

[READ MORE NEWS >](#)



UPCOMING EVENTS

Thursday, 26 May 2016

Outreach and Advocacy SIG Meeting: Faculty Outreach Strategies - Success Stories and Uphill Battles

[View Event Details »](#)

Thursday, 26 May 2016

METRO SIG Meeting: Willa Sibilla Will Co. New York

Articles ▾

FOUR RECORDINGS COVER CURRENT ASPECTS OF WEB ARCHIVING

Four webinars recorded as part of the Web Archiving Webinar series have been posted to YouTube for the enjoyment of potential web archivists around the world.

This series was co-produced by METRO's Web Archiving SIG, currently convened by Sumitra Duncan, Alex Thurman, and Anna Perricci.

Recordings include:

Web Archiving: An Overview

Featuring Karl-Rainer Blumenthal and Sumitra Duncan
Hosted by Anna Perricci

Web Archiving: Building Collections

Featuring Lori Donovan, Michael Shallcross, Dallas Pillen
Hosted by Sumitra Duncan

Web Archiving: Description and Access

Featuring Alex Thurman and Lily Pregill
Hosted by Sumitra Duncan

So You Want to Build a Web Archive...

Featuring Jefferson Bailey and Dragan Espenschied
Hosted by Anna Perricci

All of these videos can be found on [METRO's YouTube Channel](#).



Like 2 people like this. Be the first of your friends.











Save Page Now


The screenshot shows the top navigation bar of the Wayback Machine website. It includes a dark header with icons for home, search, and other functions. The main content area features the 'INTERNET ARCHIVE WayBackMachine' logo, a search bar with 'http://' entered, and a 'BROWSE HISTORY' button. Below this, a large text block states '484 billion web pages saved over time.' with a 'DONATE' link. A horizontal row of ten small thumbnail images displays various web pages from the archive. The bottom section is divided into three columns: 'Tools' with a gear icon and links to 'Wayback Machine Availability API', 'WordPress Broken Link Checker', and '404 Handler for Webmasters'; 'Subscription Service' with an 'A' icon and text about digital content collection; and 'Save Page Now' with a blue 'S' icon, a search bar, a 'SAVE PAGE' button, and text explaining the service's purpose and availability.

INTERNET ARCHIVE
WayBackMachine

http://

484 billion web pages saved over time. [DONATE](#)


         

 **Tools**


[Wayback Machine Availability API](#)
Build your own tools.

[WordPress Broken Link Checker](#)
Banish broken links from your blog.

[404 Handler for Webmasters](#)
Help users get where they were going.

 **Subscription Service**

Archive-It enables you to capture, manage and search collections of digital content without any technical expertise or hosting facilities. [Visit Archive-It to build and browse the collections.](#)

 **Save Page Now**

http://

Capture a web page as it appears now for use as a trusted citation in the future.
Only available for sites that allow crawlers.

[FAQ](#) | [Contact Us](#) | [Terms of Service \(Dec 31, 2014\)](#)

Perma CC

Perma.cc ∞

[About Perma.cc](#)

[Guide](#)

[Blog](#)

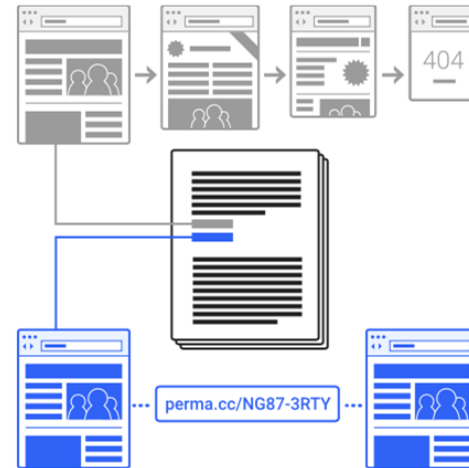
[Sign up](#)

[Log in](#)

Websites change. Perma Links don't.

Perma.cc helps scholars, journals, courts, and others create permanent records of the web sources they cite.

Perma.cc is simple, free to use, and is built and supported by libraries.



[Sign up and use Perma.cc for free](#)

[How can my library get involved?](#)

20 Years of Web Archiving

- A portion of the web has been preserved
- Increasing awareness and growing community
- High concentration of use of technology and practice
- Legal issues much better understood
- Scholarly use of web archives start to emerge
 - Maximise transparency and make base-line knowledge self-explanatory
- Focus on derivation, portability, and access
- Focus on scalable partnerships & decentralisation
- Use standards and web technology
- Full potential of web archives yet to be exploited

