

Web-Archiving: Preserving the Internet

Mark Phillips, Ph.D.

University of North Texas

UNT Libraries

February 18, 2015

The Web is the platform for communication for the
21st century.

Each year more and more content
is created digitally.

And published to the world via the Web.

Ensuring the long-term preservation and access to these resources offer new challenges and opportunities to memory organizations such as libraries, museums, and archives.

Background:

Web archiving is the process of collecting portions of the World Wide Web to ensure the information is preserved in an archive for future researchers, historians, and the public.

It involves concepts that are core to
libraries and archives.

Selection: Defining the scope of the Web for preservation.

National Domain
Federal Domain
State/Municipal Web
Subject Based
Event Based

Acquisition: Harvesting selected content by using Web crawlers such as Heritrix

**Access: Enabling interested users to accessing
Web resources from the past.**

Preservation: Ensuring the long-term access to these Web resources through systematic planning and curation.

These concepts enable memory organizations to align Web archiving with traditional collections, processes and missions

Examples:

UNT Libraries' CyberCemetery



About this Collection

Overview

At a Glance

Latest Additions

Cite This Collection

Explore Holdings

Contact Us

Items

Titles

Dates

Locations

Statistics

API

Share



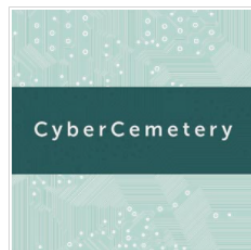
Search this Collection

Search across 121 items



Feed

CyberCemetery



About This Collection

The CyberCemetery is an archive of government web sites that have ceased operation (usually web sites of defunct government agencies and commissions that have issued a final report). This collection features a variety of topics indicative of the broad nature of government information. In particular, this collection features web sites that cover topics supporting the university's curriculum and particular program strengths.

Background

Foreseeing the potential preservation problems created by federal agencies' ventures into electronic publishing, UNT became the second depository library in the nation to join the Federal Depository Library Program Content Partnerships Program. This program attempts to ensure permanent public access to electronic federal information. As a participant, the [UNT Libraries](#), part of the Federal Depository Library Program (FDLP), were designated the host of the permanent online collection of the defunct Advisory Commission on Intergovernmental Relations (ACIR).

We've since expanded our Content Partnership with the federal government to include dozens of other defunct federal agency web sites. This electronic repository is popularly known as the CyberCemetery. In recognition of our work in this area, the UNT Libraries were designated an [Affiliated Archives](#) of the National Archives and Records Administration (NARA) in 2006. Under this agreement, the UNT Libraries will continue to preserve and provide access to the records of defunct government web sites, while the [NARA](#) will legally accession the records as part of the Archives of the United States and will join the UNT Libraries and the [Government Printing Office \(GPO\)](#) in ensuring the preservation of these valuable records. We are currently the only non-military educational institution to be designated an Affiliated Archives.

Source: Kelly, Melody Specht. [Our History: Celebrating 60 Years of Public Service](#). Accessed March 31, 2010.

How to Get Your Website in the CyberCemetery

If your agency or commission is closing, you can make sure that your website's information is preserved. For more information, see our [CyberCemetery FAQ](#).

<https://cybercemetery.unt.edu/>

Archive of Federal commissions, committees and initiatives that are no longer active.

Started in 1997 with the Advisory Commission on Intergovernmental Relations (ACIR)



Advisory Commission on Intergovernmental Relations

Welcome to The U.S. Advisory Commission on Intergovernmental Relations' Home Page

The Advisory Commission on Intergovernmental Relations (ACIR) is a permanent, independent, bipartisan intergovernmental agency established by Public Law 86-380 in 1959. As it was established, ACIR's mission is:

To strengthen the American federal system and improve the ability of federal, state, and local governments to work together cooperatively, efficiently, and effectively.

To learn more about ACIR, please click on one of the following subjects,

- [WHAT IS ACIR?](#)
- [ACIR NOW](#)
- [SERVICES, PROGRAMS, AND FUNDING](#)

scroll down this page, or
contact us directly at:
800 K St., NW
Suite 450 South Building
Washington, DC 20575

WHAT IS ACIR?

The Advisory Commission on Intergovernmental Relations (ACIR) is a permanent, independent, bipartisan agency that was established under Public Law 86-380 in 1959 to study and consider the federal government's [intergovernmental relationships](#) and the nation's intergovernmental machination.

[The Commission](#) is composed of 26 individuals who represent the interests of the federal system's intergovernmental partners in matters of intergovernmental concern, and is the only established, freestanding part of the federal system in which the views of the federal government's intergovernmental partners are openly aired, and in which difficulties and inefficiencies in the federal system's Intergovernmental Relationship's are examined.

Intergovernmental Relations:

The studies of Intergovernmental Management and Intergovernmental Relations are fields in which researchers study how different levels of government interact with one another, and attempt to define how they should interact with one another in the context of The Constitution. Intergovernmental Relationships have profound effects on self-government in America, and determining the appropriate role for federal government within the governmental system is as high on the national agenda today as it was when The Constitution was written in 1787.

As the nation's foremost repository of experience and information on intergovernmental structure, finance, process, and practice, the challenge of studying the federal government's relationships with State, Local, and Tribal governments falls to ACIR.

ACIR works with State, Local, and Tribal governments, as well as other interested parties to:

- Identify emerging intergovernmental issues, trends, and turning points;
- Stimulate thought about American federalism and intergovernmental relationships;
- Educate leaders and the public about the impacts of intergovernmental reforms; and
- Promote stronger intergovernmental communication, cooperation, and coordination as the critical basis for an effective federal system.

Currently, ACIR is the hub of the intergovernmental community and maintains relationships with many Intergovernmental organizations. Several related documents and other sources of information are:

Western Hemisphere Drug Policy Commission

[HOME](#) [DOWNLOAD REPORT](#) [READ REPORT ONLINE](#) [VIEW HEARING](#) [LEARN ABOUT](#) ▾

CHARTING A NEW PATH FORWARD
REPORT OF THE WESTERN HEMISPHERE
DRUG POLICY COMMISSION

[f Share](#) [t Tweet](#) [w Share](#) [e Email](#) [p Print](#)

Executive Summary



ABOUT **PROGRAMS** **LEARN**
partners media federal commemorations



READ THE WSCC AGENCY REPORT





UNITED STATES OF AMERICA
CYBERSPACE
SOLARIUM
COMMISSION



Introduction

The Cyberspace Solarium Commission (CSC) was [established](#) in the John S. McCain National Defense Authorization Act for Fiscal Year 2019 to "develop a consensus on a strategic approach to defending the United States in cyberspace against cyber attacks of significant consequences." The finished report was presented to the public on March 11, 2020.

Our Report

The Cyberspace Solarium Commission's proposes a strategy of layered cyber deterrence. Our report consists of over 80 recommendations to implement the strategy. These recommendations are organized into 6 pillars:

1. *Reform the U.S. Government's Structure and Organization for Cyberspace.*
2. *Strengthen Norms and Non-Military Tools.*
3. *Promote National Resilience.*
4. *Reshape the Cyber Ecosystem.*
5. *Operationalize Cybersecurity Collaboration with the Private Sector.*
6. *Preserve and Employ the Military Instrument of National Power.*

[DOWNLOAD OFFICIAL REPORT](#)

[DOWNLOAD EXECUTIVE SUMMARY](#)

[DOWNLOAD LEGISLATIVE PROPOSALS](#)



COMMISSION ON EVIDENCE-BASED POLICYMAKING

[REPORT](#) [ABOUT](#) [COMMISSIONERS](#) [MEETINGS](#) [HEARINGS](#) [LIBRARY](#) [NEWS](#) [CONTACT US](#)

The Commission on Evidence-Based Policymaking (CEP) was established to develop a strategy for increasing the availability and use of data in order to build evidence about government programs, while protecting privacy and confidentiality.

[Read More](#)

The Commission on Evidence-Based Policymaking sunsets on September 30, 2017. While this website is no longer being updated, it will be available until July 1, 2018.

For updates on related evidence-based policymaking efforts, see the Evidence-Based Policymaking Initiative of the Bipartisan Policy Center at bipartisanpolicy.org/evidence.

■ [CEP Final Report: The Promise of Evidence-Based Policymaking](#)



Latest

- News
- Meetings
- Hearings



Commission on Evidence-Based Policymaking Releases Final Report

September 06, 2017

Members of the Commission on Evidence-Based Policymaking announce the Commission's final report and recommendations.



Commission Meeting on March 13

March 01, 2017

The Commission on Evidence-Based Policymaking (CEP) will hold a public meeting on federal statistical and partnership infrastructure for evidence building.



Commission Meeting on February 24

February 22, 2017

The Commission will hold a public meeting on the role of legal standards and technology in maximizing data security and privacy.

[View All News](#)

[Privacy Policy](#)

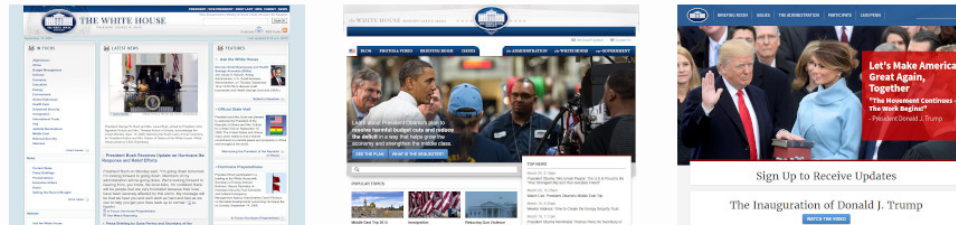
Today contains over 120 websites

UNT Libraries is an Affiliated Archives with the
U.S. National Archives and Records
Administration (NARA) because of the
CyberCemetery

End of Term Presidential Web Archives 2008, 2012, 2016, 2020

Purpose

The End of Term Web Archive captures and saves U.S. Government websites at the end of presidential administrations. The EOT has thus far preserved websites from administration changes in 2008, 2012, 2016, and 2020.



Whitehouse.gov captures from: September 15, 2008; March 21, 2013; and February 3, 2017.

Archive Scope

The End of Term Web Archive contains federal government websites (.gov, .mil, etc) in the Legislative, Executive, or Judicial branches of the government. Websites that were at risk of changing (i.e., whitehouse.gov) or disappearing altogether during government transitions were captured. Local government websites, or any other site not part of the federal government domain were out of scope.

U.S. Federal Government Domain End of Term 2020 Web Archive

For the End of Term 2020, The Library of Congress, University of North Texas Libraries, Internet Archive, Stanford University Libraries, and the U.S. Government Publishing Office (GPO) joined efforts again, this time with new partners Environmental Data & Governance Initiative (EDGI) and the National Archives and Records Administration (NARA), to preserve public United States Government websites at the conclusion of the presidential administration ending January 20, 2021. This web harvest – like its predecessors in 2008, 2012, and 2016 – was intended to document the federal government's presence on the World Wide Web during the transition of presidential administrations and to enhance the existing collections of the partner institutions. This broad comprehensive crawl of the .gov domain includes as many federal .gov sites as we could find, plus federal content in other domains (such as .mil, .com, and social media content) and FTP'd datasets.

Nominations made by individual URL for inclusion in the End of Term Presidential Harvest 2020 are available to view in the [Nomination Tool](#). URLs submitted for consideration in bulk form via files were added to a separate [bulk Nomination Tool instance](#). The files containing the bulk list URLs have also been added to a [GitHub repository](#). The entirety of the archived content is currently [being held by the Internet Archive](#).

Browse the End of Term Web Archive

- [2008-2009](#)
- [2012-2013](#)
- [2016-2017](#)
- [Browse All](#)
- [Search Full Text](#)

Collaboration between institutions throughout the United State to capture the Federal (.gov) Web at the end of each presidential term.

Good Evening



*Welcome to the
White House*

**[Social Security Administration Announces
New Online Retirement Planner](#)**



[The President & Vice President:](#)

Their accomplishments, their families, and how to send them electronic mail --



[Commonly Requested Federal Services:](#)

Direct access to Federal Services



[Interactive Citizens' Handbook:](#) Your guide to information about the Federal government



[What's New:](#)

What's happening at the White House

- [-Expanding Opportunities for America's Senior Citizens](#)**
- [-A National Call to Action to Close the Digital Divide](#)**



[White House History and Tours:](#)

Past Presidents and First Families, Art in the President's House and Tours -- **[Tour](#)**



[Site News:](#)

- Featured sections of our site
- [-White House Easter Egg Roll](#)**
- [-China Trade Relations Working](#)**



The White House

PRESIDENT GEORGE W. BUSH

President * News * Vice President * History & Tours * First Lady * Mrs. Cheney

YOUR GOVERNMENT KIDS ESPAÑOL CONTACT PRIVACY POLICY SITE MAP SEARCH



EMAIL UPDATES

SEARCH

Issues

- Economy
- Iraq
- Education
- National Security
- Homeland Security
- More Issues
- En Español

News

- Current News
- Press Briefings
- Proclamations
- Executive Orders
- Radio

Appointments

- Nominations
- Application

Offices

- Global Comm
- Freedom Corps
- Faith-Based & Community
- OMB
- More Offices

Major Speeches

- Iraq Transition
- State of the Union
- Saddam Capture
- UN Address
- National Address
- Iraqi Freedom
- National Address



PHOTOS

- Photo Essays
- Search photos by date

July 20, 2004 | Last Updated 11:00 p.m. (EST)

Secretary Snow Discusses Economy on "Ask the White House"



Treasury Secretary John Snow on Tuesday discussed the economy on "Ask the White House".

Read the Full Transcript [full story](#)

Steady Job Growth Continues in States Across the Country

June's employment data indicate that the President's economic policies are creating jobs and driving steady economic growth, but there is still more work to be done. [full story](#)

President Bush Welcomes President Lagos of Chile



President Bush and Chilean President Lagos discussed trade issues on Monday at the White House. [full story](#)

Press Briefing by Scott McClellan



Watch Monday's Press Briefing with White House Press Secretary Scott McClellan. [full story](#)

President Bush Welcomes Malaysian Prime Minister to White House



More Photos

White House photo by Paul Morse

President George W. Bush meets with Freedom Corps greeter Dean Gesme Sr. after arriving in Cedar Rapids, Iowa on July 20, 2004.

WHITE HOUSE RADIO

More ▶

Rumsfeld on the Radio

Secretary Rumsfeld talks with Martha Zoller, WDUN, Gainesville, GA

[Click here](#) for more ♦♦♦♦



ASK THE WHITE HOUSE

More ▶

Apollo 11 35th Anniversary

Discuss the 35th Anniversary of Apollo 11 tomorrow with Neil Armstrong, Michael Collins and Buzz Aldrin at 11am. Submit your question now.



WHITE HOUSE INTERACTIVE

More ▶

Email the White House...

Q: Marcus. Oaktown:





THE WHITE HOUSE

PRESIDENT GEORGE W. BUSH

Search

Podcasts RSS Feeds

November 26, 2008

Last updated 2:35 p.m. (EDT)

IN FOCUS

- [Afghanistan](#)
- [Africa](#)
- [Budget Management](#)
- [Defense](#)
- [Economy](#)
- [Education](#)
- [Energy](#)
- [Environment](#)
- [Global Diplomacy](#)
- [Health Care](#)
- [Homeland Security](#)
- [Immigration](#)
- [International Trade](#)
- [Iraq](#)
- [Judicial Nominations](#)
- [Middle East](#)
- [National Security](#)
- [Veterans](#)

[more issues](#) →

News

- [Current News](#)
- [Press Briefings](#)
- [Proclamations](#)
- [Executive Orders](#)
- [Radio](#)
- [Setting the Record Straight](#)

[more news](#) →

Interact

LATEST NEWS



[+ more photos](#)

White House Photo by Eric Draper

President George W. Bush waves to the troops at Fort Campbell, Ky., Tuesday, Nov. 25, 2008, after being introduced on stage during a Thanksgiving visit. Said the President, "Over the past seven years, folks from this base have done exactly what they were trained to do." "You have taken the battle of the terrorists overseas so we do not have to face them here in the United States. You have helped counter the hateful ideology of tyranny and terror with a more hopeful vision of justice and liberty. You're part of the great ideological struggle of our time. With the soldiers of Fort Campbell out front, the forces of freedom and liberty will prevail." White House photo by Eric Draper

President Bush Pardons "Pumpkin and Pecan" in National Thanksgiving Turkey Ceremony

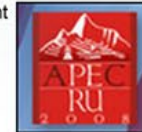


President Bush on Wednesday said, "This is my final Thanksgiving as the President. Over the past eight years, I have been given many reasons to be thankful. I'm thankful to our men and women in uniform and I am incredibly proud to have been their Commander-in-Chief. I am thankful for the armies of compassion volunteers who feed the hungry and shelter the poor. I am

FEATURES

APEC 2008

On November 21-23, 2008, President Bush attended the 16th Asia-Pacific Economic Cooperation (APEC) Leaders' Meeting in Lima, Peru.



[APEC 2008](#) →

Celebrate the 61st Annual Thanksgiving Turkey Pardon

This year marks the 61st anniversary of the National Thanksgiving Turkey presentation. Though live Thanksgiving turkeys have been presented intermittently to presidents since the Lincoln administration, the current ceremony dates to 1947, when the first National Thanksgiving Turkey was presented to President Harry Truman.



[Thanksgiving at the White House](#) →

Presidential Transition

The peaceful transfer of power from one Presidential Administration to the next is a hallmark of American democracy. With our Nation at war, our homeland targeted by terrorist adversaries, and our economy facing serious challenges, the Administration is committed to establishing and executing a transition plan that minimizes disruption, maintains continuity, and addresses the major changes in government since the 2000 transition, including the Intelligence





Get Email Updates | Contact Us

[BLOG](#) [PHOTOS & VIDEO](#) [BRIEFING ROOM](#) [ISSUES](#) [the ADMINISTRATION](#) [the WHITE HOUSE](#) [our GOVERNMENT](#)



America has never been about what can be done for us. It's about what can be done by us, together, through the hard and frustrating but necessary work of self-government. That's the principle we were founded on.

— President Barack Obama, November 7, 2012

[JOIN PRESIDENT OBAMA](#)



IN FOCUS



Helping Survivors of

TOP NEWS

November 11, 11:55am
Thank an American Hero

November 10, 5:45am
Weekly Address: Extending Middle Class Tax Cuts to Grow the Economy



BRIEFING ROOM

ISSUES

THE ADMINISTRATION

PARTICIPATE

1600 PENN

Search



Stay in the Loop

From photos to the President's schedule to policy updates, the Daily Snapshot has everything you need to know about a day at the White House.

START NOW

POPULAR TOPICS



Net Neutrality

President Obama is asking the



2015 State of The Union

Did you miss this year's speech?



Immigration Action

The President is taking

TOP NEWS

February 12, 4:20pm
The Clay Hunt Act: What the President Just Signed

February 12, 5:32pm
President Obama, a Selfie Stick, and HealthCare.gov

February 11, 8:28pm
The Authorization of Military Force Against ISIL Terrorists: What You Need to Know



BRIEFING ROOM

ISSUES

THE ADMINISTRATION

PARTICIPATE

1600 PENN



“Believe in yourselves. Believe in your future.
And believe, once more, in America.”

- President Donald J. Trump

Sign Up to Receive Updates

FEATURED CONTENT



President Trump Praises Law Enforcement While Delivering Remarks on MS-13

President Trump gave remarks to local, State, and Federal law enforcement about the challenges of, and the Administration's success regarding, MS-13.

[READ MORE](#)



HELP IS HERE

WE PASSED THE AMERICAN RESCUE PLAN



- See How The American Rescue Plan Helps People Like You
- Submit Your Story About The American Rescue Plan
- Learn How Working Americans Will Receive Their Direct Payments



Selection, harvesting, archiving distributed to meet local collection needs.

Centralized access

UNT Libraries has acquired archive for 2008 and 2012 resulting in a 16TB and 30TB dataset

2016 - 159TB (HTTP) 150TB (FTP)
2020 - 300TB

Working to move a copy of EOT archives to Open Data on AWS

Based on the work by Common Crawl
Designed for computational access to datasets.

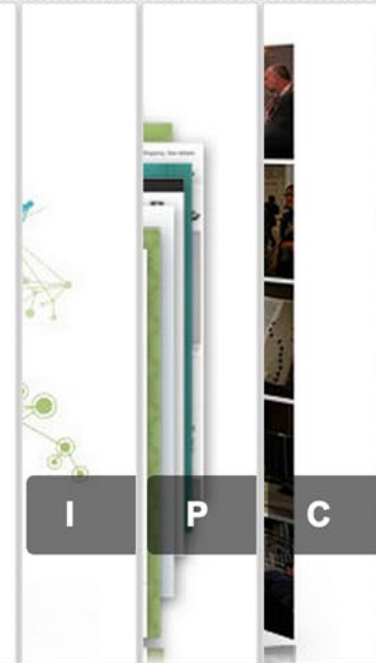
Using Web Archives in scholarship, teaching and research

Collaborating Internationally

International Internet Preservation Consortium (netpreserve.org)

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

[Learn More About IIPC](#)

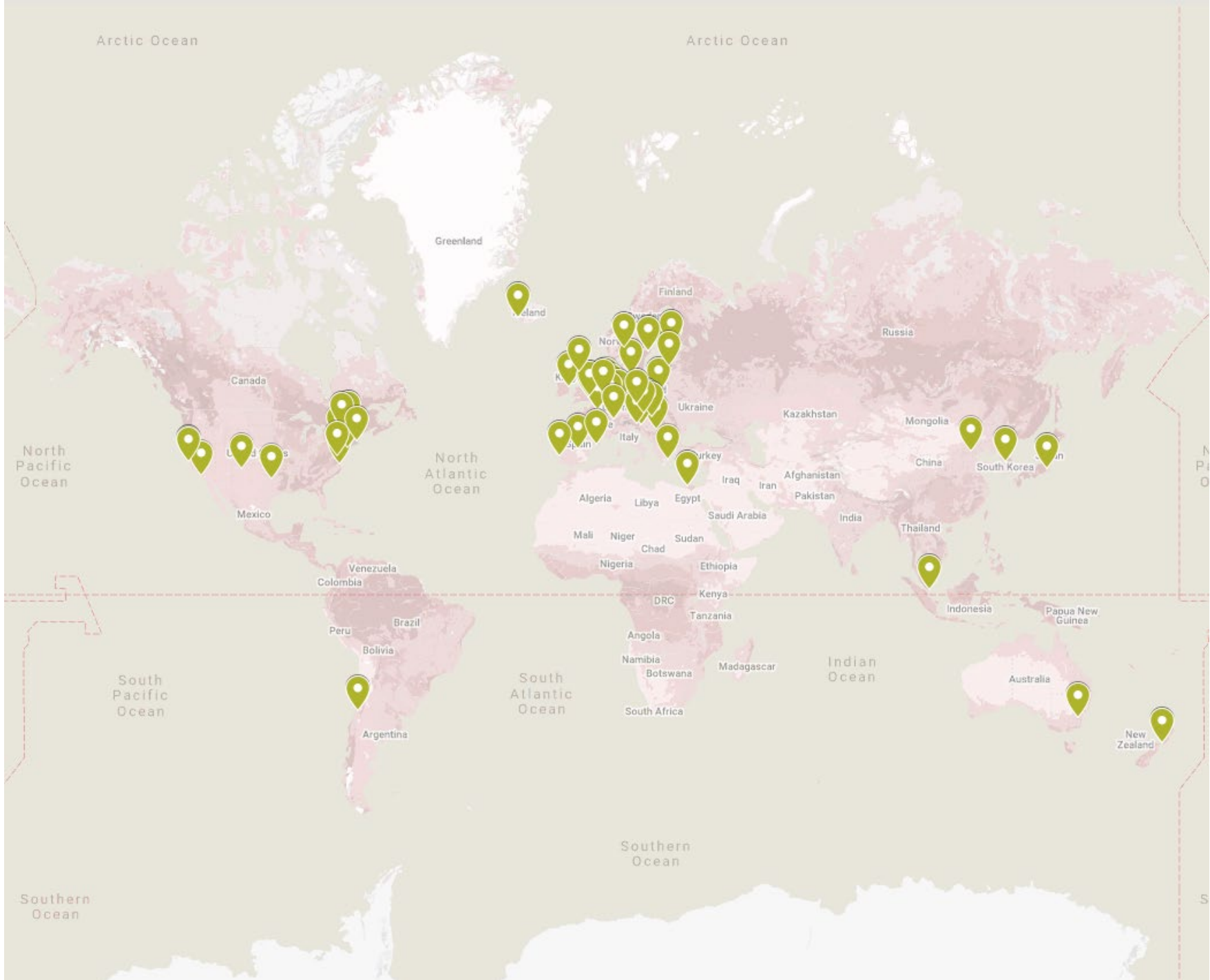


WHY ARCHIVE THE WEB?

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



55 member institutions



Members from 25 countries: North America,
Europe, Asia, Africa

National Libraries
National Archives
Research Libraries
Government Agencies
Private Companies

Common goal to preserve the Web

Collaboratively develop tools for the community such as the Heritrix Web crawler and the Open Wayback system for replay of Web archives.

Search the history of over 452 billion pages on the Internet.



http://www.



Web

Texts

Video

Audio

Software

Image

Search

Anonymous



EXIT BETA

FEEDBACK



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



7.6M



1.9M



2.3M



96K



948K



148K



119K

Universal Access to Knowledge

GO

Announcements

Locking the Web Open, a Call for a Distributed Web

Knight Foundation to Support Toolsets for Building Libraries Together

The New Yorker: The Cobweb—Can the Internet be archived?

SEE MORE

Top Collections at the Archive

Lead work on the formalization of the WARC File Format (ISO 28500:2009)

Web archiving is an important component in preserving a national heritage for future generations.

Thank you.

Mark Phillips

mark.phillips@unt.edu

@vphill