

THE LIBRARY OF CONGRESS: DIGITAL PRESERVATION VIDEO SERIES

Web Archiving

The World Wide Web has dramatically changed how we get and use information. An amazingly detailed record of our era is now only available on the web. But content on the web changes all the time. If we don't save that content before it disappears, a major part of our history will be lost.

Why would web content disappear? Here's an example: When elections end, candidates have little incentive to maintain their campaign web sites. These sites have unique value for students, historians and the general public.

The Library of Congress and its partners are capturing, preserving and providing permanent access to web-based information. Web archiving supports the Library's mission, to build a universal collection of knowledge and creativity.

There are challenges to web archiving. Here are four examples.

- First, we need to define the scope of what to collect. It's impractical to collect everything on the web, so the Library focuses on significant themes or events.
- A second challenge is legal. The Library has to operate within existing intellectual property laws and request permissions from website owners.
- A third challenge is technological. Web sites are dynamic, ranging in size from one page to millions of pages. Sites can include many complex elements, such as computer code, graphics, text, audio and video. The Library uses specialized tools to download copies of websites.
- A fourth challenge is access. After the Library downloads a copy of a website, the website is reassembled and stored in a digital archive and made available for researchers. The Library captures websites at regular intervals. This provides researchers with a record of how websites change over time.

Other organizations are also archiving websites. The Library collaborates with the International Internet Preservation Consortium, which consists of national libraries and other organizations from around the world.

So, in summary, the Library of Congress is working to provide permanent access to web content of historical importance.

It selects websites for collection, requests permissions from the website owners, addresses the technology of collecting websites and preserves the websites and makes them available.

To learn more about web archiving at the Library of Congress and elsewhere, please visit loc.gov/webarchiving/.