

Welcome to the September 11 Web Archive

Our aim in creating this collection of Web materials was to preserve the Web expressions of individual people, groups, the press and institutions from around the world, in the aftermath of the attacks in the U.S. on September 11, 2001. We hope the archive provides resources for many kinds of reflection on the meanings of these events.

Starting in the hours just after the attacks, until December 1, 2001, we collected Web materials that reflect responses to them, from as many sources as possible. People from around the world contributed URLs to be included in this collection, and we are grateful for each submission.

We have catalogued approximately 2,500 sites related to the September 11 attacks. This catalogued collection is now available through the [Minerva Project of the Library of Congress](#). We continue to explore new ways of making sense of these archived materials. We welcome your feedback about how you use the archive, and suggestions for improving access to the archived materials. Please send email to us at september11@loc.gov.

Developing The September 11 Web Archive

In the immediate aftermath of the terrorist attacks on September 11, 2001, producers of Web sites around the world employed information production and hyperlinking capabilities in attempts to provide information, features, news, services, reactions, as well as virtual memorials, for various audiences around the world. Between September 11 and December 1, 2001, scholars affiliated with WebArchivist.org worked with staff at the Library of Congress, the [Internet Archive](#), and volunteers from around the world to identify URLs that were likely to be relevant to the question of how Web site producers were reacting to the events of September 11. We identified eight categories of site producers that we expected to be responding on the Web to the attacks. These included: news organizations; governmental agencies; charity/relief groups; educational organizations; religious groups; advocacy groups; commercial organizations and individual citizens. We conducted systematic searches for URLs produced by these sets of actors, and followed links to find other URLs with relevant content. In most cases, the salient feature of these sites was content referring to the attacks and/or their aftermath. In some cases, the absence or removal of such content was salient.

Our initial efforts identified nearly 30,000 different "sites." Each of the identified sites was archived on a daily basis until December 1, 2001. The objective of the archiving activity was to preserve not only the bits and the content, but also the experiential dimensions of this rapidly emerging Web sphere. By capturing pages and sites in their hyperlinked context, the archiving tools preserved not just the collection of Web pages, but an interlinked Web sphere -- characterized and bounded by a shared object orientation or reference point, in this case, the September 11 attacks. The entire archive consists of 5 TB of Web materials.

WebArchivist.org, in work funded by the Library of Congress and the **Pew Internet & American Life Project**, developed the software to search the portion of the collection that is fully catalogued. This activity included developing methods to link both human and machine-generated metadata to objects (sites, pages, and page components) in the archive, adding significant value for users of the archive, including researchers, journalists and citizens. In addition, WebArchivist.org researchers have documented the types of online action made possible on more than 600 archived pages.

Scholars, journalists and citizens are increasingly recognizing the value of Web sphere archives that allow us to retrospectively assess the impact and importance of the Web in critical social and political events. This project represents one of the first attempts to dynamically capture, on a large scale, a rapidly emerging Web sphere for future analysis. The nature of the September 11 terrorist attacks presented several challenges to established methods of Web archiving. The scale of the response on the Web to the attacks was unprecedented within the relatively short history of the Web, and resulted in a very large Web sphere. In addition, the inability to anticipate the event precluded any preparations for site identification and archiving strategies prior to the event itself, and required the development of a fully dynamic site identification process.