

About the International Internet Preservation Consortium

In July 2003 the national libraries of Australia, Canada, Denmark, Finland, France, Iceland, Italy, Norway, Sweden, The British Library (UK), The Library of Congress (USA) and the Internet Archive (USA) acknowledged the importance of international collaboration for preserving Internet content for future generations. This group of 12 institutions chartered the IIPC to fund and participate in projects and working groups to accomplish the Consortium's goals. In 2007, membership was opened to other libraries, archives, and other cultural heritage institutions interested in this important endeavor. Membership inquiries should be addressed to membership@netpreserve.org.

2007 Members

Asia

- National Library Board, Singapore

Australia/Oceania

- National Library of Australia
- National Library of New Zealand

Europe

- Library of Catalonia
- National Library of France: *Coordinating Institution and Technical Officer*
- British Library (U.K.)
- German National Library
- European Archive Foundation
- Hanzo Archives Ltd. (U.K.)
- The National Library of Finland
- National Library of the Netherlands
- National Library of Sweden

- National and University Library of Iceland
- National and University Library, Slovenia
- National Library of the Czech Republic
- The National Library of Norway
- National Archives (U.K.)
- National Library of Scotland
- Netarchive.dk (The Royal Library and the State and University Library, Denmark)
- Swiss National Library

North America

- California Digital Library
- Internet Archive
- Library and Archives Canada
- Library of Congress: *Communications and Membership*
- Library of Virginia
- United States Government Printing Office

Goals

- To enable the collection, preservation and long-term access of a rich body of Internet content from around the world.
- To foster the development and use of common tools, techniques and standards for the creation of international archives.
- To be a strong international advocate for initiatives and legislation that encourage the collection, preservation and access to Internet content.
- To encourage and support libraries, archives, museums and cultural heritage institutions everywhere to address Internet content collecting and preservation.

For more information about the IIPC :
<http://www.netpreserve.org>

Technical Work Plan for 2007-2009

Working Groups

To achieve its mission, the consortium has set up dedicated committees consisting of members of some of the participating libraries working on specific topics and providing the consortium with various deliverables. A technical committee supervises and runs projects which have an impact on the overall Web archiving framework and technical architecture. It guarantees convergence and consistency of standards and practices in areas such as harvesting, access and preservation. The 2007 chartered working groups are Harvesting, Access, Preservation, and Standards.

Harvesting

In 2007, the Harvesting Working Group's primary focus is the development of a smart crawler. Other areas of focus include:

- Development and support of the WARC file format
- Best practices
- Feature requests for crawler
- Harvesting the deep web
- Harvesting video and streaming media

Access

The Access Working Group will focus on initiatives, procedures and tools required to provide immediate access and to preserve the future access to Internet material in a Web archive. Focus areas include:

- Defining User Requirements to improve existing access tools (Wayback Machine, WERA)
- Testing full text indexing using NutchWax
- Defining requirements for user authentication/authorization/access controls
- Access tools for the analysis of the content of the archived internet material

Preservation

The IIPC Preservation Working Group is looking at the extent to which existing digital preservation standards, practices and approaches can be applied to the management of Web archives. Over the past decade, there has been great attention paid to the processes of capturing online resources, as a necessary step in their preservation; however, work on maintaining accessibility for the long term remains reasonably undeveloped. At the same time, many approaches have been proposed and implemented for other kinds of digital collections. The Preservation Working Group aims to understand and report on how such approaches might be used with Web archives, as well as the special characteristics of Web archives that might require new approaches.

Standards

IIPC work on standards is dependent on the directions and priorities of the three other working groups. In the short term, the IIPC is working through the ISO standard adoption process. Future investigations may involve other standards, APIs, metadata, and metrics.

For more information about the IIPC :
<http://www.netpreserve.org>