



International Internet Preservation Consortium 2008 Member Profile Survey Results

Author: Abbie Grotke,
International Internet Preservation Consortium
Communications Officer
membership@netpreserve.org
netpreserve.org

Date of issue: December 16, 2008

Status: Final

Table of Contents

1	About the International Internet Preservation Consortium	3
2	About the Survey	3
3	Responders	4
4	Responses	5
4.1.1	Membership Type and Maturity	5
4.1.2	Web Archiving Staff	6
4.1.3	Scope, Selection Guidelines, and Metrics	8
4.1.4	Crawl and Curator Tools	10
4.1.5	Access and Description	11
4.1.6	Legal Issues and Policies	13

Document Control

<i>Issue</i>	<i>Date of issue</i>	<i>Comments</i>
1	December 16, 2008	Public Release

1 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. The initial agreement was in effect for three years, during which time the membership was limited to the charter institutions. Since then, membership has expanded to include additional libraries, archives, museums and cultural heritage institutions involved in Web archiving. The goals of the consortium are:

- 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 consortium and to see a complete list of members, visit <http://www.netpreserve.org>.

2 About the Survey

The IIPC Member Profile Survey opened to members on March 7, 2008, with preliminary results presented during the membership report at an April 2008 General Assembly meeting in Canberra, Australia. Additional responses to the survey were received after the April briefing and were recorded in this report.

This summary of survey results published here for a wider web archiving and digital library community presents answers to portions of the original survey. The survey was

divided into two sections, "Part 1: About You and Your Web Archiving Activities" and "Part 2: About Your IIPC Participation: Your Contributions and Expectations." This report shares portions of results from Part 1; some questions in Part 1 and all questions in Part 2 have been kept confidential to IIPC members.

Many of the questions allowed for free-form answers, and for confidentiality reasons, those detailed responses for the most part have not been included in this public report.

Readers should be aware that the allowance of comments as responses created some skewed results in the "response percent" and "response count" columns, since "other" and "comments" were allowed in the same field.

3 Responders

Of 39 IIPC members in 2008, 35 responded to the survey, with 32 completing all questions (91.4%).

4 Responses

4.1.1 Membership Type and Maturity

4.1.1.1 Type of Organization

Answer Options	Response Percent	Response Count
National Library	68.6%	24
Other type of Library (see other below)	11.4%	4
National Archives	2.9%	1
University/College	14.3%	5
Museum	0.0%	0
Service Provider	2.9%	1
Other: <ul style="list-style-type: none"> Statutory Board that manages National Library, Reference Libraries, Academic Libraries and Public Libraries Government Agency or Department National Library, National Archive, Portrait Gallery Non Profit Foundation or Library State Library and Archives Research Organization or Center University Library 	31.4%	11
<i>answered question</i>		35
<i>skipped question</i>		0

4.1.1.2 What is the status or maturity of the web archiving activity in your institution?

Answer Options	Response Percent	Response Count
----------------	------------------	----------------

Experimental – we’re trying to figure out what to do and how to do it	20.0%	7
Project – we’ve got special funding or a time-limited project	22.9%	8
Operational but still experimenting – our staff is full time on web archiving but we are still finding our way	40.0%	14
Fully Operational - we are fully integrated into the organization’s activities	17.1%	6
Web archiving is our primary business or a dedicated and operating branch of our activity	11.4%	4
Other	14.3%	5
<i>answered question</i>		35
<i>skipped question</i>		0

4.1.2 Web Archiving Staff

4.1.2.1 Describe the location of the web archiving unit, team and staff within your organization.

Answer Options	Response Percent	Response Count
All of your web archiving staff are in one division or section.	25.7%	9
Your web archiving staff are located in multiple areas of your organization.	57.1%	20
Other	20.0%	7
<i>answered question</i>		35
<i>skipped question</i>		0

4.1.2.2 Identify how many staff working on web archiving in what areas of your organization.

IIPC Member Profile Survey Results 2008												
Identify how many staff working on web archiving in what areas of your organization. For each area listed, select how many staff - and whether full time, part time, or occasionally involved. Explain further in comments if needed.												
Full Time												
Answer Options	1	2	3	4	5	6	7	8	9	10	More*	Response Count
Collections Unit or Division	3	0	1	2	1	0	0	0	0	0	0	7
Digital Library	2	1	1	1	1	0	1	0	0	0	1	8
Information Technology Office	4	1	1	0	0	0	0	0	0	0	2	8
Legal Deposit Office	2	0	1	1	0	0	0	0	0	0	0	4
Cataloging Office	0	1	2	1	0	1	0	0	0	0	0	5
Other	3	0	0	0	0	0	0	0	0	0	1	4
Part Time												
Answer Options	1	2	3	4	5	6	7	8	9	10	More*	Response Count
Collections Unit or Division	4	2	0	0	0	0	0	1	0	0	1	8
Digital Library	7	1	0	2	1	0	0	0	0	0	1	12
Information Technology Office	8	3	3	0	0	0	0	0	0	0	1	15
Legal Deposit Office	5	2	1	0	0	0	0	0	0	0	0	8
Cataloging Office	4	1	0	0	0	0	0	0	0	0	0	5
Other	0	3	0	0	0	1	1	0	0	0	0	5
Occasional Support or Participation												
Answer Options	1	2	3	4	5	6	7	8	9	10	More*	Response Count
Collections Unit or Division	0	3	0	0	0	1	0	0	0	1	3	8
Digital Library	3	1	0	0	0	0	0	0	0	0	1	5
Information Technology Office	6	6	1	0	0	0	0	0	0	0	0	13
Legal Deposit Office	1	2	0	0	0	0	0	0	0	0	0	3
Cataloging Office	1	1	0	1	0	0	0	0	0	0	0	3
Other	1	0	0	0	0	0	0	0	0	0	1	2
											answered question	34
											skipped question	1

4.1.3 Scope, Selection Guidelines, and Metrics

4.1.3.1 What is the scope of your archiving?

Answer Options	Response Percent	Response Count
Broad or bulk crawls	20.0%	7
Entire national domain	40.0%	14
Regional domain	17.1%	6
Thematic/event/selective archiving	94.3%	33
Other, or comments	42.9%	15
<i>answered question</i>		35
<i>skipped question</i>		0

4.1.3.2 Do you have published selection guidelines or collection development guidelines?

Answer Options	Response Percent	Response Count
No, we don't have written guidelines	29.4%	10
No, they are for internal use only	20.6%	7
Yes, but not on a public website	23.5%	8
Yes.	26.5%	9
<i>answered question</i>		34
<i>skipped question</i>		1

We asked in a follow up question whether the guidelines were in English or in other languages. Fourteen members have their guidelines in English, while 13 have guidelines in their native languages, but not translated to English at the time of the survey.

4.1.3.3 What statistics do you use when reporting to management or others in your organization about web archiving activities?

Answer Options	Response Percent	Response Count
URIs: pages	15.2%	5
URIs: documents	27.3%	9
URIs: items	33.3%	11
Hosts	21.2%	7
Records	9.1%	3
Crawls	30.3%	10
Seeds	27.3%	9
ARC/WARC	9.1%	3
Size (Terabytes/Gigabytes, etc.)	63.6%	21
Web sites or other internet resources selected	54.5%	18
Web sites or other internet resources described/cataloged	24.2%	8
Other, or comments	45.5%	15
	<i>answered question</i>	33
	<i>skipped question</i>	2

NOTE: Four members commented that they had not yet determined their web archiving metrics or statistics.

4.1.4 Crawl and Curator Tools

4.1.4.1 What crawler or web archiving service do you use?

Answer Options	Response Percent	Response Count
Crawling in-house, using: <ul style="list-style-type: none"> • Heritrix (various versions mentioned): <ul style="list-style-type: none"> ○ 1.10, 10.2 ○ 1.12, 1.12.1, 1.12.2 ○ 1.6 ○ 2.x (“most recent”) • Teleport Pro • Heritrix via curator tools: <ul style="list-style-type: none"> ○ WCT 1.2.7 ○ WCT 1.3.0 ○ Heritrix 1.12.1 + NetarchiveSuite • HTTrack/Pandora • Archive-IT • SrceHarvester 	79.4%	27
Contract out crawling services (such as Internet Archive or European Archive)	17.6%	6
Use services such as Hanzo or Archive-IT	8.8%	3
Not applicable	8.8%	3
Other, or comments	70.6%	24
<i>answered question</i>		34
<i>skipped question</i>		1

4.1.4.2 What curator tool, if any, are you using?

Answer Options	Response Percent	Response Count
NetArchive	6.1%	2
Pandas	6.1%	2
Web Curator Tool	18.2%	6
We're testing still but undecided	18.2%	6
Not applicable	18.2%	6
Other, or comments	60.6%	20
<i>answered question</i>		33
<i>skipped question</i>		2

4.1.5 Access and Description

4.1.5.1 What access tool are you using?

Answer Options	Response Percent	Response Count
Wayback Machine	53.3%	16
WERA	13.3%	4
Other, or comments	66.7%	20
<i>answered question</i>		30
<i>skipped question</i>		5

Note: Five members commented that they had not yet determined or decided on their access tools.

4.1.5.2 Have you done full-text indexing of any of your collections?

Open-Ended Response	Response Count
Full text indexing is being tested and implemented across the IIPC, using various tools and methods: <ul style="list-style-type: none"> • NutchWAX • WERA • Accela Bizsearch • Home-grown search engines • via Archive-It 	29
<i>answered question</i>	29
<i>skipped question</i>	6

Nine members responded no to this question. Two responded with “not yet,” but indicated they had plans to do so.

4.1.5.3 Do you catalog or describe your archives?

Answer Options	Response Percent	Response Count
Site-level or Item-level cataloging	58.1%	18
Collection-level cataloging	22.6%	7
No cataloging	25.8%	8
Other, or comments	32.3%	10
<i>answered question</i>		31
<i>skipped question</i>		4

4.1.5.4 What metadata format do you use for description of your resources?

Answer Options	Response Percent	Response Count
Dublin Core	46.4%	13
MARC	50.0%	14
MODS	14.3%	4
Other, or comments	50.0%	14
<i>answered question</i>		28
<i>skipped question</i>		7

4.1.6 Legal Issues and Policies

4.1.6.1 Do you have any legal authority or laws in your country related to web archiving?

Answer Options	Response Percent	Response Count
Law enacted or passed (allowing you to?)*	15.6%	5
Law expected (when?)*	12.5%	4
Lobbying in progress	15.6%	5
Not Applicable	6.3%	2
Other, or comments	50.0%	16
<i>answered question</i>		32
<i>skipped question</i>		3

4.1.6.2 Do you have to seek permission or negotiate rights to...

Answer Options	Response Percent	Response Count
Harvest or Crawl	32.3%	10
Allow researcher access	29.0%	9
Not Applicable	9.7%	3
Other, or comments	64.5%	20
<i>answered question</i>		31
<i>skipped question</i>		4

4.1.6.3 What, if any, access restrictions do you have?

Answer Options	Response Percent	Response Count
Everything is in a dark archive but Web archiving or other staff can see the content	16.1%	5
Completely dark archive, no one can have access	3.2%	1
Access is open to researchers onsite at your institution	22.6%	7
Access is open to researchers onsite and offsite (via the Internet) for anyone to see	29.0%	9
Other, or comments	67.7%	21
<i>answered question</i>		31
<i>skipped question</i>		4