LIBER QUARTERLY 19 (2), December 2009 – ISSN: 1435-5205. P140–165
http://liber.library.uu.nl/
Igitur, Utrecht Publishing & Archiving Services

Facilitating Searches in Multiple Bibliographical Databases: Metadata Harvesting Service Providers

Mangala Anil Hirwade

Lecturer, Department of Library & Information Science,
RTM Nagpur University, Nagpur,
hirwade2004@indiatimes.com

Mohini T. Bherwani

Librarian, Shri Binzani City College, Umrer Road, Nagpur,
mohinibherwani@yahoo.co.in

Abstract

A metadata harvester is a software package that reads data from servers, writes it to databases, implements various kinds of searches, and writes HTML files to display the results. In this paper sixty metadata harvesting service providers have been studied. The study reviewed metadata generation, preservation and harvesting, and various technical issues arising at these stages.

Key words: metadata; metadata interoperability; harvesting; OAI-PMH; service providers; data providers

1. Introduction

In the digital environment new methodologies of information management and access, coupled with advancements in digital information systems, have transformed to a great extent the ways and means of information management. Metadata, the systematic arrangement of data elements, aids the identification and location of information resources, thereby facilitating improved access to them. However, there exists unpredictability in terms of the

availability, accessibility and authenticity of digital objects. Many search mechanisms retrieve a plethora of information resources, but the majority lack effectiveness and comprehensiveness.

2. Objectives

The objectives of the present study are

- to discuss the importance of metadata harvesting service providers for the next generation library interface.
- to trace various metadata harvesting service providers.
- to study the technical details, features, metadata generation and preservation tools, server requirements, metadata elements and user support system used by those metadata harvesting service providers.

3. Methodology

The study focuses on the current status of sixty metadata harvesting service providers. The paper is largely based on a review of the literature, both online and print. The data for this paper was downloaded from the official websites of these metadata harvesting service providers during July–August 2009.

4. Review of Literature

4.1 Defining Metadata

Metadata is structured information that describes, explains, locates or otherwise makes it easier to retrieve, use or manage an information resource. Metadata is often called data about data or information about information (UKOLN website)

4.2 Interoperability

Interoperability in relation to metadata is search interoperability, or the ability to perform a search over diverse sets of metadata records and obtain meaningful results. Different individuals or organisations may have created metadata according to the same scheme or they may have applied of multiple schemes, as different metadata schemes serve distinct needs and audiences. Complementary schemes can be used to describe the same resource for multiple purposes and to serve a number of user groups (Baker, 2009).

There is a need to interrelate sources and types of information with different formats, data structures and description standards. Using metadata to record data about information sources allows an initial assessment of compatibility and provides an avenue for merging information or for exchanging information between systems. Interoperability is the ability of multiple systems with different hardware and software platforms, data structures and interfaces to exchange data with minimal loss of context and functionality (ALCTS/CCS Committee on Cataloging, 2000).

4.3 Metadata Harvesting

A metadata harvesting service harvests or indexes metadata from open access initiative (OAI) compliant archives or repositories through harvesting software that supports a protocol known as the Open Access Initiative Protocol for Metadata Harvesting (OAI-PMH). It is designed for better sharing and retrieval of e-prints residing in distributed archives, allow resources to be found by relevant criteria, identifying resources, bringing similar resources together and giving location of information (Hodge, 2003).

Harvesting refers to the activity of searching for and collecting metadata from Open Archives Initiative (OAI) Institutional Repositories (IR's) whose content is indexed and posted for open use from a World Wide Web server. An OAI harvester is software that performs the job of regularly 'visiting' open access databases that have informed the harvester of their existence. The harvested metadata is accrued in a database that can then be searched. The harvester's creator decides what services to provide on top of this data, for example, searching and cross-linking. The harvester can be set to harvest only metadata on a specific subject, from a select group of data providers, or from all available open access databases. The harvested metadata is archived and preserved. The Institutional Repositories commit to upgrade accessibility as technology changes. The OAI/PHM protocol is an international standard of classification fields for any item that is shared in an OAI archive such as author, content description, abstract, type of file, and other 'tags' that classify content in ways that can be stored and retrieved from a data base server (Coleman, 2008).

4.4 Metadata Harvesting Protocol

As the term denotes, a metadata harvesting protocol sets rules or guidelines for harvesting metadata.

In order to facilitate metadata harvesting, there ought to be some agreement on aspects such as: the transport protocol (HTTP or FTP etc.), the metdata format (Dublin Core, MARC, etc.); metadata quality assurance (mandatory element set, name and subject conventions, etc.) and intellectual property and usage rights.

The OAI protocol for metadata harvesting provides an application-independent interoperability framework which can be used by a variety of communities who are engaged in publishing content on the web. It provides a set of rules that defines the communication between systems such as FTP or HTTP on the internet. That is why even though the protocol actually uses HTTP as a transport mechanism between digital libraries, it is popularly known as the 'HTTP of digital libraries'.

There are two classes of players in the OAI-PMH framework: *data providers*, which administer systems that support the OAI-PMH as a means of exposing metadata, and *service providers*, which use metadata harvested via the OAI-PMH as a basis for building value-added services.

The protocol based on HTTP and XML was developed with the objective to ensure interoperability between e-print repositories only. Later, in version 1.0/1.1, all document-like digital objects were brought within its purview, and finally the latest version 2.0 supports all kinds of digital resources.

It must be emphasised that OAI-PMH is not a search engine or a search tool or a database. It only provides a set of rules for moving the metadata (not the content) of the digital resource from one repository to another. The content remains in the source repository. A repository can act both as a service provider or harvester and data provider, or only as a service provider or data provider. The protocol is not restricted to supporting simple metadata (unqualified Dublin Core), but can support any metadata schema which can be provided in an XML format (Munshi, 2009).

5. Analysis and Interpretation

5.1 Metadata harvesting service providers

A total of sixty metadata harvesting service providers were traced during the study; they were grouped as shown in Table 1.

Table 1. Basic details of metadata service providers.

Sr. No.	Name of the Harvesting Service	Abbreviation used	URL	Name of the parent body
1	Search Digital Library	SDL	http://drtc.isibang.ac.in/sdl/	DRTC, Bangalore
2	Scientific journal publishing in India	SJPI	http://144.16.72.144/harvester/	NCSI, IISc
3	Search engine for engineering digital repositories	SEED	http://eprints.iitd.ac.in/seed/	IIT, Delhi
4	Open J Gate	Open J Gate	www.openj-.gate.com	Informatics India Ltd
5	Open Index Initiative	Open Index	http://oii.igidr.ac.in	Indira Gandhi Institute of Development Research Reserve Bank of India, Government of India
6	Knowledge Harvester	Knowledge Harvester	http://61.16.154.195/harvester	INSA, India
7	Cross Archive Search Service for Indian Repositories	CASSIR	http://casin.ncsi.iisc.ernet.in/oai/; http://ardb4.ncsi.iisc-.ernet.in/oai/	National Centre for Science Information, Indian Institute of Science, Bangalore
8	Prototype digital archive of Indian aerospace research	P-DAINAR	www.ncsi.iisc.ernet.in	National Center for Science Information
9	IWF Metadata harvester	IWF	http://savannah.nongnu.org/projects	IWF Wissen & Media gGmbH
10	Latin America Open Archive Portal	LAOAP	http://lanic.utexas.edu	LARRP & LANI
11	Latin American knowledge harvester	LAKH	http://lakh.unm.edu	University of New Mexico
12	International Association of Aquatic & Marine Science Libraries & Information Centers	IAMSLIC	www.iamslic.org	IAMSLIC
13	Archives and museum informatics	Archimuse	www.archimuse.com	David Bearman & Jennifer Trant
14	D-Space metadata harvester	D-Space	www.dspace.org/introduction intro-faculty.html-9k	MIT Libra & Hewlett Packard Laboratories
15	Canadian association of research libraries	CARL	www.david mattison.cal/wordpress/?	Canadian association of research libraries

Table 1. (Continued)

Sr. No.	Name of the Harvesting Service	Abbreviation used	URL	Name of the parent body
16	Public knowledge project harvester	PKP	www.pkp.sfu.ca	PKP group
17	Ibreo-American Scientific & technical educational consortium	ISTEC	www.istec.org	Ibreo-American Scientific & Technical educational consortium
18	Networked computer science technical research library	NCSTRL	www.ncstrl.org	Networked computer science technical research library
19	Art, Design, Architecture & Media	ADAM	www.adam.ac.uk	ADAM group of UK
20	Association of college and research libraries	ACRL	www.ala.org	American library association
21	Resource Organisation & discovery in subject based services	ROADS	www.ukoln.ac.uk/roads/harvester	Electronic libraries programme of UK
22	NTRS-NASA Technical Reports server	NTRS	http://ntrs.nasa.gov/	NASA
23	Community Research & Development Information	CORDIS	www.cordis.europa.eu/data	Spanish Council Presidency
24	Rexahn Pharmaceutical	RNN	www.rareextreme.com/ferums/index.php	Rexahn Pharmaceutical
25	Joint information system committee	JISC	www.jisc.ac.uk	JISC board, UK
26	Networked digital libraries of theses & dissertations	NDLTD	www.alcme.oclc.org	Online computer library center
27	SOLINET	SOLINET	www.solinet.edu	University of Alberta
28	Economic & social data services	ESDS	www.esds.ac.uk/about.asp	Economic & social research Council of UK
29	British university's film & video counseling	UKOLN	www.bufvc.ac.uk	British universities film & video counseling
30	METALIS	METALIS	www.metalis.org	Not known
31	UIUC Digital Gateway to Cultural Heritage Materials	DGCHM	www.uiuc.dgchm.org	UNIMP, UK
32	Scholarly publishing & academic resources coalition	SPARC	www.cni.org	Association of research libraries, UK

Table 1. (Continued)

Sr. No.	Name of the Harvesting Service	Abbreviation used	URL	Name of the parent body
33	MetaArchive.org	MetaArchive.org	www.metaarchieve.com	Division of telecommunication Research, UK
34	Australian government national archives of Australia	AGLS	www.naa.gov.au/default.htm	Australian Government
35	Global change master directory	GCMD	www.gcmd.gsfc.nasa.gov.html	NASA
36	Aristocrat Industries Incorporation	ARII	www.ncsc.online.org	ARII, Australia
37	Deutsche Initiative for Netzwerkin Formation	DINI	www.dini.de/document/DINI	University of Kassel
38	OARiNZ Harvester	OARiNZ	www.oarinz.ac.nz	Christchurch Polytechnic Institute of Technology (CPIT)
39	California digital library	CDL	www.cdl.org	University of California
40	Digital publishing Technology Center	D Pubs	www.dpubs.calgary	University of Calgary
41	OAIster	OAIster	www.oaister.umich.edu	University of Michigan
42	American South.org	American south	http://www.americansouth.org/	---------
43	ARC: A cross archive service	ARC	www.arc.cs.odu.edu	---------
44	ARCHON	ARCHON	www.archon.org	Not given
45	The Directory of Open Access Repositories	DOAR	http://www.opendoar.org	University of Nottingham
46	SAIL-eprints	SAIL eprints	http://eprints.bo.cnr.it	---------
47	Sheetmusic Service	Sheetmusic	http:digital.library.ucla.edu/sheetmusic/librarian?	Sheetmusic Consortium
48	Registry of open access repositories	ROAR	http://archives.eprints.org/	---------
49	Celestial open archives gateway	Celestial	http://celestial.eprints.org	---------
50	Experimental OAI Registry at UIUC	Experimental, UIUC	http://gita.grainger.ui.ac.edu/registry	---------
51	Directory of mathematics Preprint and e-Print servers	Mathematics e print	http://www.ams.org/global-preprints	---------

Table 1. (Continued)

Sr. No.	Name of the Harvesting Service	Abbreviation used	URL	Name of the parent body
52	Digital commons	Digital commons	http://digitalcommons.library.tmc.edu	---------
53	Eprints Archive	Eprints Archive	http://www.eprints.org/software/archives/	---------
54	Digital Academic repositories	DARE	http://www.darenet.nl/en	---------
55	Open Language Archives Community	Open language	http://www.language.archive.org	Open Language Archives Community
56	CERN document server	CERN	http://cdsweb.cern.ch/	---------
57	TORII, Digital gateway to cultural heritage materials	TORII	http://torii.digi.edu/	---------
58	Scirus, the web search engine for scientific information	Scirus	www.scirus.ac.uk/	---------
59	Resource discovery network	RDN	http://www.rdn.ac.uk/	---------
60	CYCLADES	CYCLADES	http://nergal.grainger.uiuc.edu.	OCLC

5.2 Technical Details

The technical details of the metadata harvesting service providers were analysed as shown in Table 2.

Table 2. Technical details of metadata harvesting service providers.

Sr. No.	Name of the harvesting service	Country	Standard Adopted	Software Used	No. of Repositories being harvested	Subject
1	SDL	India	Dublin core	PKP	27	Library & Information Sciwence
2	SJPI	India	Dublin core	PKP	13	Science
3	SEED	India	Dublin core	PKP	4	Engineering
4	Open J Gate	India	Dublin core	PKP	4300+	Multidisciplinary
5	Open Index	India	Self developed	Self developed	16	Multidisciplinary
6	Knowledge Harvester	India	Dublin core	PKP	3	Multidisciplinary
7	CASSIR	India	Dublin core	PKP	18	Science & Technology
8	P-DAINAR	India	Dublin core	PKP	5	Aerospace
9	IWF	US	Dublin core	PKP	17	Multidisciplinary
10	LAOAP	US	TEI	E prints	23	Multidisciplinary
11	LAKH	US	Dublin core	PKP	24	Science
12	IAMSLIC	US	Dublin core	Dspace	27	Multidisciplinary
13	Archimuse	US	METS, CDWA Lite, MPEG 7	Fedora	600	Multidisciplinary
14	D-Space	US	Dublin core, © Metadata	Dspace	254	Multidisciplinary
15	CARL	US	METS	CDSware	28	Multidisciplinary
16	PKP	US	Dublin core, NISO MIX, Darwin core	E prints	7	Multidisciplinary
17	ISTEC	US	EAD, LOM, CIDOC CRM	PKP, Dspace	286	Multidisciplinary
18	NCSTRL	US	MODS, METS, Darwin core	Dspace	51	Science & Technology
19	ADAM	US	MODS, EAD, IPTC	Fedora	600	Computer science
20	ACRL	Germany	Dublin core	PKP	2500	Multidisciplinary
21	ROADS	US	TEI	CDSware	39	Multidisciplinary
22	NTRS	UK	EAD, MODS,IPTC	ROADS	37	Multidisciplinary
23	CORDIS	UK	EAD	Fedora	31	Multidisciplinary
24	RNN	UK	Dublin core	E prints	3	Multidisciplinary
25	JISC	UK	TEI, AACR 2, MARC 21 XML	CDSware	39	Medicines
26	NDLTD	UK	TEI, Darwin core	Dspace	67	Multidisciplinary
27	SOLINET	Germany	Dublin core, EAD, TEI	CDSware	10	History

Table 2. (Continued)

Sr. No.	Name of the harvesting service	Country	Standard Adopted	Software Used	No. of Repositories being harvested	Subject
28	ESDS	UK	TEI, DDI	PKP	95	Multidisciplinary
29	UKOLN	UK	METS, GEM, AGLS	Fedora	4	Social & Economics
30	METALIS	UK	TEI, CDWA Lite, GEM	CDSware	10	Photography
31	DGCHM	UK	EAD, TEI, MODS	E prints	23	Multidisciplinary
32	SPARC	UK	EAD, ONIX, Darwin core	CDSware	300	Multidisciplinary
33	MetaArchive.org	UK	EAD, AGLS, LOM	PKP	83	Multidisciplinary
34	AGLS	UK	TEI	CDSware	71	Science & Tech.
35	GCMD	Australia	Dublin core	CDSware	10	Multidisciplinary
36	ARII	Australia	TEI	PKP	45	Multidisciplinary
37	DINI	Australia	EAD	Eprints	06	Multidisciplinary
38	OARiNZ	New Zealand	Dublin core	PKP	11	Multidisciplinary
39	CDL	Netherlands	Dublin core	E prints	11	Nuclear science
40	D Pubs	Indonesia	EAD	E prints	61	Multidisciplinary
41	OAIster	Indonesia	EAD	Fedora	1155	Multidisciplinary
42	American south	Atlanta	Dublin core	PKP	86	Multidisciplinary
43	ARC	Caribbean	Dublin core	Dspace	679	General, education
44	ARCHON	US	Dublin core	PKP	32	Physics
45	DOAR	Nottingham	Dublin core	Not known	1473	Multidisciplinary
46	SAIL eprints	US	Dublin core	Not known	53	Science
47	Sheetmusic	UK	ONIX, IPTC, NISO MIX	Dspace	300	Music
48	ROAR	US	MPEG 7, TEI, MODS	Dspace	1418	Multidisciplinary
49	Celestial	Italy	Dublin core, Darwin core, AGLS	PKP	46	Science & Technology
50	Experimental, UIUC	US	LOM, EAD, MODS	HTML	340	Multidisciplinary
51	Mathematics e print	US	MARC 21 XML, AGLS, Darwin core	VITAL	23	Mathematics
52	Digital commons	US	DDI, METS, NISO MIX	Digital Commons	209	Science & Technology
53	Eprints Archive	US	CIDOC CRM, IPTC, © Metadata	Dspace	679	General, education
54	DARE	UK	METS, LOM ONIX	PKP	Not mentioned	Education
55	Open language	Belgium	EAD, MODS, NISO MIX	Dspace	9	Multidisciplinary
56	CERN	US	MPEG 7, IPTC, LOM	CDSware	Not mentioned	Science & Technology
57	TORII	UK	CDWA Lite, LOM	Digital Commons	9	Multidisciplinary

Table 2. (Continued)

Sr. No.	Name of the harvesting service	Country	Standard Adopted	Software Used	No. of Repositories being harvested	Subject
58	Scirus	UK	EAD, LOM	Dspace	358	Multidisciplinary, education
59	RDN	US	Dublin core, CIDOC CRM, TEI	Dspace	7	Multidisciplinary
60	CYCLADES	US	METS, DDI, GILS	Eprints	13	Multidisciplinary

From the Table 2 it can be seen that the United States is the leading country when it comes to metadata harvesting service providers: it has 22 service providers (36.66%), followed by the United Kingdom which has 16 (26.66%). Only eight providers (13.33%) were established in India. It can also be observed from Table 2 that Dublin Core is the most popular metadata standard used by metadata harvesting service providers. 25 harvesters (41.66%) use Dublin Core, 21.66% use EAD. AACR2 is used by only one service provider.

PKP is the most popular software used by 31.66% of the service providers, followed by Dspace. 15% of the service providers use CDS ware. 37 service providers (61.66%) are multidisciplinary, 6 (10%) are science and technology-specific, 1 (1.66%) harvests metadata in the field of library and information science, and 4 (6.66%) in science and education.

5.3 Analysis of Metadata Generation of Harvesters

Table 3 presents details on metdata generation:

Table 3. Metadata Generation of Harvesters.

		Type of Metadata			Creator of metadata		Type of creation		Tools used for Metadata creation			
Sr. No.	Name of the Harvester	Descriptive	Structural	Administrative	Technical Staff	Originator of the resource	Machine Generated	Human Generated	Template	Markup Tools	Extraction Tools	Conversion Tools
1	SDL	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
2	SJPI	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
3	SEED	No	Yes	No	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
4	Open J Gate	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
5	Open Index	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 3. (Continued)

Sr. No.	Name of the Harvester	Type of Metadata			Creator of metadata		Type of creation		Tools used for Metadata creation			
		Descriptive	Structural	Administrative	Technical Staff	Originator of the resource	Machine Generated	Human Generated	Template	Markup Tools	Extraction Tools	Conversion Tools
6	Knowledge Harvester	Yes	No	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
7	CASSIR	Yes	Yes	Yes	Yes	No	No	Yes	Yes	No	Yes	Yes
8	P-DAINAR	Yes	Yes	Yes	Yes	No	Yes	No	No	Yes	Yes	No
9	IWF	No	Yes	No	Yes	No	Yes	No	Yes	Yes	Yes	Yes
10	LAOAP	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
11	LAKH	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No	Yes
12	IAMSLIC	Yes	No	Yes	No	No	Yes	Yes	Yes	Yes	Yes	Yes
13	Archimuse	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
14	D-Space	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
15	CARL	Yes	No	Yes	No	No	Yes	No	No	Yes	Yes	No
16	PKP	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
17	ISTEC	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
18	NCSTRL	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	No	No	Yes
19	ADAM	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	Yes
20	ACRL	No	Yes	No	Yes	No	Yes	Yes	Yes	No	Yes	No
21	ROADS	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
22	NTRS	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
23	CORDIS	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
24	RNN	Yes	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
25	JISC	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No	Yes
26	NDLTD	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	No
27	SOLINET	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	No	Yes	Yes
28	ESDS	No	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
29	UKOLN	Yes	Yes	Yes	No	No	Yes	Yes	No	Yes	Yes	Yes
30	METALIS	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
31	DGCHM	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	No
32	SPARC	No	Yes	Yes	Yes	No	No	No	Yes	Yes	No	Yes
33	MetaArchive.org	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes
34	AGLS	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
35	GCMD	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
36	ARII	Yes	No	Yes	Yes	No	Yes	No	Yes	No	Yes	Yes
37	DINI	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
38	OARiNZ	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	Yes	No
39	CDL	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	No	Yes
40	D Pubs	Yes	Yes	Yes	Yes	No	Yes	Yes	No	Yes	Yes	Yes
41	OAIster	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
42	American south	No	Yes	Yes	Yes	No	Yes	No	Yes	Yes	Yes	No
43	ARC	Yes	Yes	Yes	No	Yes	Yes	No	No	Yes	Yes	Yes
44	ARCHON	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No
45	DOAR	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
46	SAIL eprints	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	No	Yes
47	Sheetmusic	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 3. (Continued)

Sr. No.	Name of the Harvester	Type of Metadata			Creator of metadata		Type of creation		Tools used for Metadata creation			
		Descriptive	Structural	Administrative	Technical Staff	Originator of the resource	Machine Generated	Human Generated	Template	Markup Tools	Extraction Tools	Conversion Tools
48	ROAR	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	Yes
49	Celestial	Yes	Yes	Yes	Yes		No	Yes	Yes	No	Yes	No
50	Experimental, UIUC	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
51	Mathematics eprint	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No
52	Digital commons	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No	Yes	Yes
53	Eprints Archive	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	No	Yes
54	DARE	Yes	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
55	Open language	No	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
56	CERN	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
57	TORII	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes
58	Scirus	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No
59	RDN	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
60	CYCLADES	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes

From Table 3 it can be observed that 48 harvesters (80%) produce descriptive metadata, 53 (88.33%) structural metadata and 51 (85%) administrative metadata. In 51 cases (85%) the metadata are generated **automatically and in** 45 cases (75%) they are produced manually. 49 Harvesters (81.66%) use templates for metadata creation, 49 (81.66%) use mark-up tools, 52 (86.66%) use extraction tools and 49 (81.66%) use conversion tools for metadata generation.

5.4 Analysis of Metadata Generation Tools

Table 4 shows the tools used for metadata generation:

Table 4. Metadata Generation Tools.

Sr. No.	Name of the harvesting service	New record creation	Record			
			Edition	Validation	Withdrawal	Redeposition
1	SDL	Yes	Yes	Yes	Yes	Yes
2	SJPI	Yes	Yes	Yes	Yes	Yes
3	SEED	Yes	Yes	Yes	Yes	Yes
4	Open J Gate	Yes	Yes	Yes	Yes	Yes
5	Open Index	Yes	Yes	Yes	Yes	Yes

Table 4. (Continued)

Sr. No.	Name of the harvesting service	New record creation	Record			
			Edition	Validation	Withdrawal	Redeposition
6	Knowledge Harvester	Yes	Yes	Yes	Yes	No
7	CASSIR	Yes	Yes	Yes	Yes	Yes
8	P-DAINAR	Yes	No	Yes	Yes	Yes
9	IWF	Yes	Yes	Yes	Yes	Yes
10	LAOAP	Yes	Yes	Yes	Yes	Yes
11	LAKH	Yes	Yes	Yes	Yes	Yes
12	IAMSLIC	Yes	Yes	Yes	Yes	Yes
13	Archimuse	Yes	Yes	Yes	Yes	Yes
14	D-Space	Yes	No	Yes	Yes	Yes
15	CARL	Yes	Yes	Yes	Yes	No
16	PKP	Yes	Yes	Yes	Yes	Yes
17	ISTEC	Yes	Yes	Yes	Yes	Yes
18	NCSTRL	Yes	Yes	Yes	Yes	Yes
19	ADAM	Yes	No	Yes	Yes	Yes
20	ACRL	Yes	Yes	Yes	Yes	Yes
21	ROADS	Yes	Yes	Yes	Yes	Yes
22	NTRS	Yes	Yes	Yes	Yes	Yes
23	CORDIS	Yes	No	Yes	Yes	No
24	RNN	Yes	Yes	Yes	Yes	Yes
25	JISC	Yes	Yes	Yes	Yes	Yes
26	NDLTD	Yes	Yes	Yes	Yes	Yes
27	SOLINET	Yes	Yes	Yes	Yes	Yes
28	ESDS	Yes	Yes	Yes	Yes	Yes
29	UKOLN	Yes	Yes	Yes	Yes	Yes
30	METALIS	Yes	No	Yes	Yes	No
31	DGCHM	Yes	No	Yes	Yes	Yes
32	SPARC	Yes	Yes	Yes	Yes	Yes
33	MetaArchive.org	Yes	Yes	Yes	Yes	Yes
34	AGLS	Yes	Yes	Yes	Yes	Yes
35	GCMD	Yes	Yes	Yes	Yes	No
36	ARII	Yes	Yes	Yes	Yes	Yes
37	DINI	Yes	No	Yes	Yes	Yes
38	OARiNZ	Yes	Yes	Yes	Yes	No
39	CDL	Yes	Yes	Yes	Yes	Yes
40	D Pubs	Yes	No	Yes	Yes	No
41	OAIster	Yes	No	Yes	Yes	Yes
42	American south	Yes	Yes	Yes	Yes	Yes
43	ARC	Yes	No	Yes	Yes	No
44	ARCHON	Yes	Yes	Yes	Yes	Yes
45	DOAR	Yes	Yes	Yes	Yes	No
46	SAIL eprints	Yes	No	Yes	Yes	Yes
47	Sheetmusic	Yes	Yes	Yes	Yes	Yes
48	ROAR	Yes	No	Yes	Yes	Yes
49	Celestial	Yes	Yes	Yes	Yes	Yes

Table 4. (Continued)

Sr. No.	Name of the harvesting service	New record creation	Record			
			Edition	Validation	Withdrawal	Redeposition
50	Experimental, UIUC	Yes	No	Yes	Yes	No
51	Mathematics e print	Yes	No	Yes	Yes	Yes
52	Digital commons	Yes	Yes	Yes	Yes	No
53	Eprints Archive	Yes	Yes	Yes	Yes	Yes
54	DARE	Yes	No	Yes	Yes	Yes
55	Open language	Yes	Yes	Yes	Yes	Yes
56	CERN	Yes	Yes	Yes	Yes	Yes
57	TORII	Yes	No	Yes	Yes	No
58	Scirus	Yes	Yes	Yes	Yes	No
59	RDN	Yes	Yes	Yes	Yes	Yes
60	CYCLADES	Yes	No	Yes	Yes	No

From Table 4 it can be observed that all the 60 harvesters (100%) facilitate new record generation, 43 (71.67%) have facilities for editing records and 53 (88.33%) provide record redeposition provision. All the service providers supply record validation and withdrawal services.

5.5 Metadata Preservation Tools

These are the metadata preservation tools used (Table 5):

Table 5. Metadata Preservation Tools.

Sr. No.	Name of the harvesting service	Provenance	Authenticity	Preservation activity	Technical environment	Rights management
1	SDL	Yes	Yes	Yes	Yes	Yes
2	SJPI	Yes	Yes	Yes	Yes	Yes
3	SEED	No	Yes	Yes	No	Yes
4	Open J Gate	Yes	Yes	Yes	Yes	Yes
5	Open Index	Yes	Yes	Yes	Yes	Yes
6	Knowledge Harvester	Yes	Yes	Yes	Yes	Yes
7	CASSIR	Yes	Yes	Yes	Yes	Yes
8	P-DAINAR	Yes	No	Yes	Yes	Yes
9	IWF	Yes	Yes	Yes	No	No
10	LAOAP	Yes	Yes	No	Yes	Yes
11	LAKH	Yes	Yes	Yes	Yes	Yes
12	IAMSLIC	Yes	Yes	Yes	Yes	Yes
13	Archimuse	Yes	Yes	Yes	Yes	Yes
14	D-Space	Yes	Yes	Yes	Yes	Yes

Table 5. (Continued)

Sr. No.	Name of the harvesting service	Provenance	Authenticity	Preservation activity	Technical environment	Rights management
15	CARL	Yes	Yes	Yes	Yes	Yes
16	PKP	No	Yes	Yes	Yes	Yes
17	ISTEC	Yes	Yes	Yes	Yes	No
18	NCSTRL	Yes	Yes	No	Yes	Yes
19	ADAM	Yes	Yes	Yes	Yes	Yes
20	ACRL	Yes	Yes	Yes	No	Yes
21	ROADS	No	No	Yes	Yes	Yes
22	NTRS	Yes	Yes	Yes	Yes	Yes
23	CORDIS	Yes	Yes	Yes	Yes	Yes
24	RNN	Yes	Yes	Yes	Yes	No
25	JISC	Yes	Yes	No	Yes	Yes
26	NDLTD	No	Yes	Yes	Yes	Yes
27	SOLINET	Yes	No	Yes	Yes	Yes
28	ESDS	Yes	Yes	Yes	No	Yes
29	UKOLN	Yes	Yes	Yes	Yes	Yes
30	METALIS	Yes	Yes	Yes	Yes	Yes
31	DGCHM	Yes	Yes	Yes	Yes	Yes
32	SPARC	Yes	Yes	Yes	Yes	Yes
33	MetaArchive. org	Yes	Yes	No	Yes	Yes
34	AGLS	No	Yes	Yes	Yes	Yes
35	GCMD	Yes	No	Yes	Yes	No
36	ARII	Yes	Yes	Yes	Yes	Yes
37	DINI	No	Yes	Yes	No	Yes
38	OARiNZ	Yes	Yes	No	Yes	No
39	CDL	Yes	No	Yes	Yes	Yes
40	D Pubs	Yes	Yes	Yes	Yes	Yes
41	OAIster	Yes	Yes	Yes	No	No
42	American south	Yes	Yes	Yes	Yes	Yes
43	ARC	Yes	Yes	No	Yes	Yes
44	ARCHON	Yes	Yes	Yes	Yes	Yes
45	DOAR	Yes	Yes	Yes	No	Yes
46	SAIL eprints	Yes	No	Yes	Yes	No
47	Sheetmusic	Yes	Yes	No	Yes	Yes
48	ROAR	No	Yes	Yes	Yes	Yes
49	Celestial	Yes	Yes	Yes	Yes	Yes
50	Experimental, UIUC	Yes	Yes	Yes	No	Yes
51	Mathematics e print	Yes	No	Yes	Yes	Yes
52	Digital commons	No	Yes	Yes	Yes	No
53	Eprints Archive	Yes	Yes	No	Yes	Yes
54	DARE	Yes	Yes	Yes	No	Yes

Table 5. (Continued)

Sr. No.	Name of the harvesting service	Provenance	Authenticity	Preservation activity	Technical environment	Rights management
55	Open language	No	Yes	Yes	Yes	Yes
56	CERN	Yes	No	Yes	Yes	Yes
57	TORII	Yes	Yes	No	Yes	No
58	Scirus	No	Yes	Yes	No	Yes
59	RDN	Yes	Yes	Yes	Yes	Yes
60	CYCLADES	Yes	Yes	Yes	Yes	Yes

Table 5 shows that 49 (81.66%) consider provenance for metadata preservation, 52 (86.66%) authenticity, 51 (85%) preservation activity, 50 (83.33%) technical environment and 51 (85%) consider rights management for metadata preservation.

5.6 Analysis of Metadata Elements

Table 6 shows the metadata elements used by the metadata harvesting service providers.

Table 6. Use of metadata elements.

Sr. No.	Name of the Repository	Title	Creator	Subject	Description	Publisher	Contributor	Date
1	SDL	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2	SJPI	Yes	Yes	No	Yes	No	No	Yes
3	SEED	Yes	Yes	Yes	Yes	Yes	Yes	Yes
4	Open J Gate	No	Yes	No	No	No	Yes	Yes
5	Open Index	Yes	Yes	Yes	Yes	Yes	Yes	Yes
6	Knowledge Harvester	Yes	Yes	Yes	Yes	Yes	Yes	No
7	CASSIR	Yes	Yes	No	Yes	No	Yes	Yes
8	P-DAINAR	Yes	Yes	No	Yes	Yes	Yes	Yes
9	IWF	Yes	Yes	Yes	Yes	Yes	Yes	Yes
10	LAOAP	Yes	Yes	Yes	Yes	Yes	Yes	Yes
11	LAKH	Yes	Yes	No	No	Yes	No	No
12	IAMSLIC	Yes	Yes	Yes	Yes	Yes	Yes	Yes
13	Archimuse	No	Yes	Yes	Yes	Yes	Yes	Yes
14	D-Space	Yes	Yes	Yes	Yes	No	Yes	Yes
15	CARL	Yes	Yes	Yes	Yes	Yes	Yes	No
16	PKP	Yes	Yes	Yes	No	Yes	No	Yes
17	ISTEC	Yes	Yes	Yes	Yes	Yes	Yes	Yes
18	NCSTRL	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 6. (Continued)

Sr. No.	Name of the Repository	Title	Creator	Subject	Description	Publisher	Contributor	Date
19	ADAM	Yes	Yes	No	No	Yes	Yes	No
20	ACRL	Yes	Yes	Yes	Yes	Yes	Yes	Yes
21	ROADS	Yes	Yes	Yes	Yes	Yes	Yes	Yes
22	NTRS	Yes	No	Yes	No	No	Yes	Yes
23	CORDIS	Yes	Yes	Yes	Yes	Yes	No	Yes
24	RNN	No	Yes	Yes	Yes	Yes	Yes	Yes
25	JISC	Yes	Yes	No	Yes	Yes	Yes	Yes
26	NDLTD	Yes	Yes	Yes	Yes	No	Yes	No
27	SOLINET	Yes	Yes	Yes	Yes	Yes	Yes	Yes
28	ESDS	Yes	Yes	Yes	Yes	Yes	Yes	Yes
29	UKOLN	No	No	Yes	Yes	Yes	Yes	Yes
30	METALIS	Yes	Yes	Yes	Yes	Yes	Yes	Yes
31	DGCHM	Yes	Yes	Yes	Yes	Yes	Yes	Yes
32	SPARC	Yes	Yes	No	Yes	No	Yes	Yes
33	MetaArchive. org	No	Yes	Yes	No	Yes	Yes	Yes
34	AGLS	Yes	Yes	Yes	Yes	Yes	Yes	Yes
35	GCMD	Yes	Yes	Yes	Yes	Yes	Yes	Yes
36	ARII	Yes	No	Yes	Yes	Yes	Yes	Yes
37	DINI	No	Yes	Yes	No	No	No	Yes
38	OARiNZ	Yes	Yes	Yes	No	Yes	Yes	Yes
39	CDL	Yes	Yes	Yes	No	Yes	Yes	No
40	D Pubs	Yes	No	Yes	Yes	Yes	Yes	Yes
41	OAIster	Yes	Yes	No	Yes	Yes	Yes	Yes
42	American south	Yes	Yes	Yes	Yes	Yes	No	Yes
43	ARC	No	Yes	Yes	Yes	Yes	Yes	Yes
44	ARCHON	Yes	No	No	No	Yes	Yes	No
45	DOAR	Yes	Yes	Yes	Yes	Yes	Yes	Yes
46	SAIL eprints	Yes	Yes	Yes	Yes	Yes	Yes	Yes
47	Sheetmusic	Yes	Yes	Yes	Yes	Yes	Yes	Yes
48	ROAR	Yes	Yes	No	No	Yes	Yes	Yes
49	Celestial	No	Yes	Yes	Yes	Yes	No	Yes
50	Experimental, UIUC	Yes	Yes	No	Yes	Yes	No	Yes
51	Mathematics e print	Yes	Yes	Yes	Yes	Yes	Yes	No
52	Digital commons	Yes	Yes	Yes	No	No	Yes	Yes
53	Eprints Archive	Yes	No	Yes	Yes	Yes	Yes	Yes
54	DARE	Yes	Yes	Yes	Yes	Yes	Yes	No
55	Open language	Yes	Yes	No	Yes	No	Yes	Yes
56	CERN	No	Yes	Yes	Yes	Yes	Yes	Yes
57	TORII	Yes	No	Yes	No	Yes	No	Yes
58	Scirus	Yes	Yes	Yes	Yes	Yes	Yes	Yes
59	RDN	Yes	Yes	Yes	Yes	Yes	Yes	Yes
60	CYCLADES	No	Yes	Yes	No	Yes	Yes	Yes

50 harvesters (83.33%) use title, 53 (88.33%) creator, 47 (78.33%) subject, 46 (76.66%) description, 50 (83.33%) publisher, 51 (85%) contributor and 51 (85%) use date as metadata element.

5.7 User Support System

Metadata harvesting service providers maintain a strong user support system, which helps the user to navigate with ease and retrieve relevant documents. The user support systems are described in Table 7.

Table 7. User support system.

Sr. No.	Name of the Repository	Navigation links	Browse Interface	Simple search interface	Advanced search interface	Result Set Processing	Sorting Fields	Hit Frequency	Display Record	Usage Statistics	Duplicate record detection	Standardization of Archive names	Cross Citation	Alerting services	Archive search service
1	SDL	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
2	SJPI	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
3	SEED	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes	No
4	Open J Gate	Yes	No	Yes	No	No	No	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
5	Open Index	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
6	Knowledge Harvester	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
7	CASSIR	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No	No
8	P-DAINAR	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
9	IWF	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
10	LAOAP	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
11	LAKH	No	Yes	Yes	Yes	Yes	Yes	No	Yes	No	No	Yes	No	Yes	No
12	IAMSLIC	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
13	Archimuse	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
14	D-Space	Yes	No	Yes	Yes	No	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
15	CARL	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
16	PKP	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
17	ISTEC	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
18	NCSTRL	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
19	ADAM	No	Yes	Yes	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
20	ACRL	Yes	No	Yes	Yes	Yes	Yes	No	Yes	Yes	No	Yes	No	Yes	No
21	ROADS	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
22	NTRS	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
23	CORDIS	No	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 7. (Continued)

Sr. No.	Name of the Repository	Navigation links	Browse Interface	Simple search interface	Advanced search interface	Result Set Processing	Sorting Fields	Hit Frequency	Display Record	Usage Statistics	Duplicate record detection	Standardization of Archive names	Cross Citation	Alerting services	Archive search service
24	RNN	Yes	Yes	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
25	JISC	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes	No	Yes
26	NDLTD	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes	No	Yes	Yes
27	SOLINET	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
28	ESDS	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
29	UKOLN	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
30	METALIS	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
31	DGCHM	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
32	SPARC	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
33	MetaArchive.org	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
34	AGLS	Yes	No	Yes	No	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes	Yes
35	GCMD	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
36	ARII	No	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	No
37	DINI	Yes	Yes	Yes	Yes	Yes	No	No	Yes	No	Yes	Yes	Yes	Yes	Yes
38	OARiNZ	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
39	CDL	Yes	No	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
40	D Pubs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
41	OAIster	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
42	American south	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes
43	ARC	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
44	ARCHON	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes	No	Yes
45	DOAR	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
46	SAIL eprints	No	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
47	Sheetmusic	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
48	ROAR	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
49	Celestial	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
50	Experimental, UIUC	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
51	Mathematics eprint	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
52	Digital commons	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
53	Eprints Archive	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
54	DARE	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
55	Open language	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
56	CERN	No	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
57	TORII	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes	Yes	No	Yes	Yes
58	Scirus	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	No
59	RDN	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
60	CYCLADES	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Fifty service providers (83.33%) provide navigation links, simple & advanced Search; alerting services are provided by 52 harvesters (86.66%). Duplicate record deletion is the feature of 88.33% of the harvesters.

5.8 Display Options

The display options provided by each harvester are shown in Table 8.

Table 8. Display options.

Sr. No.	Name of the Repository	Title	Author	Date stamp	Discovery date	Archives	Subject	Hit Frequency	Citations hits
1	SDL	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2	SJPI	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
3	SEED	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
4	Open J Gate	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
5	Open Index	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
6	Knowledge Harvester	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
7	CASSIR	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
8	P-DAINAR	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
9	IWF	Yes	Yes	No	Yes	Yes	Yes	No	Yes
10	LAOAP	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
11	LAKH	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
12	IAMSLIC	Yes	No	No	Yes	Yes	Yes	Yes	No
13	Archimuse	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
14	D-Space	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
15	CARL	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
16	PKP	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
17	ISTEC	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
18	NCSTRL	Yes	Yes	Yes	Yes	No	No	Yes	Yes
19	ADAM	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
20	ACRL	Yes	No	Yes	Yes	Yes	Yes	Yes	No
21	ROADS	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
22	NTRS	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
23	CORDIS	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
24	RNN	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
25	JISC	Yes	Yes	Yes	Yes	No	Yes	No	Yes
26	NDLTD	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
27	SOLINET	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
28	ESDS	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
29	UKOLN	Yes	No	Yes	Yes	Yes	No	Yes	Yes
30	METALIS	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
31	DGCHM	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 8. (Continued)

Sr. No.	Name of the Repository	Title	Author	Date stamp	Discovery date	Archives	Subject	Hit Frequency	Citations hits
32	SPARC	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
33	MetaArchive.org	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
34	AGLS	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
35	GCMD	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
36	ARII	No	Yes	Yes	No	Yes	Yes	Yes	Yes
37	DINI	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
38	OARiNZ	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
39	CDL	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
40	D Pubs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
41	OAIster	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
42	American south	Yes	No	Yes	Yes	Yes	No	Yes	Yes
43	ARC	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
44	ARCHON	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
45	DOAR	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
46	SAIL eprints	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
47	Sheetmusic	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
48	ROAR	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
49	Celestial	No	Yes	Yes	No	Yes	Yes	Yes	Yes
50	Experimental, UIUC	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
51	Mathematics e print	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
52	Digital commons	Yes	Yes	Yes	Yes	No	Yes	No	Yes
53	Eprints Archive	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
54	DARE	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
55	Open language	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
56	CERN	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
57	TORII	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
58	Scirus	No	Yes	Yes	Yes	Yes	No	Yes	Yes
59	RDN	Yes	No	Yes	No	Yes	Yes	Yes	No
60	CYCLADES	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

55 service providers (91.66%) display the title metadata element, 54 (90%) the author, 55 (91.66%) the date stamp, 55 (91.66%) the discovery date, 56 (93.33%) the name of the archive, 53 (88.33%) the subject of the content, 56 (93.33%) the hit frequency and 53 (88.33%) citation hits.

5.9 Error Elements

The error elements are shown in Table 9.

Table 9. Error elements.

Sr. No.	Name of the Repository	Bad argument	Bad resumption token	Bad verb	Cannot disseminate Format	Id does not exist	No records match	No metadata formats	No set Hierarch
1	SDL	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2	SJPI	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
3	SEED	Yes	Yes	No	Yes	No	Yes	Yes	Yes
4	Open J Gate	No	Yes	Yes	Yes	Yes	Yes	No	Yes
5	Open Index	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
6	Knowledge Harvester	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
7	CASSIR	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
8	P-DAINAR	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
9	IWF	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
10	LAOAP	Yes	No	Yes	Yes	Yes	Yes	Yes	No
11	LAKH	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
12	IAMSLIC	Yes	Yes	Yes	No	Yes	Yes	Yes	No
13	Archimuse	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
14	D-Space	Yes	Yes	Yes	Yes	No	Yes	No	Yes
15	CARL	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
16	PKP	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
17	ISTEC	No	Yes	Yes	Yes	Yes	Yes	Yes	No
18	NCSTRL	No	Yes	Yes	Yes	Yes	No	Yes	Yes
19	ADAM	Yes	No	No	Yes	Yes	Yes	Yes	Yes
20	ACRL	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
21	ROADS	Yes	Yes	Yes	No	No	Yes	Yes	No
22	NTRS	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
23	CORDIS	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
24	RNN	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
25	JISC	Yes	Yes	No	Yes	Yes	No	Yes	Yes
26	NDLTD	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
27	SOLINET	No	Yes	Yes	Yes	Yes	Yes	No	No
28	ESDS	Yes	No	Yes	Yes	No	Yes	Yes	Yes
29	UKOLN	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
30	METALIS	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
31	DGCHM	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 9. (Continued)

Sr. No.	Name of the Repository	Bad argument	Bad resumption token	Bad verb	Cannot disseminate Format	Id does not exist	No records match	No metadata formats	No set Hierarch
32	SPARC	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
33	MetaArchive.org	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
34	AGLS	Yes	No	Yes	Yes	Yes	No	Yes	Yes
35	GCMD	Yes	Yes	No	Yes	Yes	Yes	Yes	No
36	ARII	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
37	DINI	Yes	Yes	Yes	Yes	No	Yes	No	Yes
38	OARiNZ	No	No	Yes	No	Yes	Yes	Yes	Yes
39	CDL	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
40	D Pubs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
41	OAIster	Yes	Yes	No	Yes	Yes	No	Yes	No
42	American south	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
43	ARC	Yes	No	Yes	Yes	No	Yes	No	Yes
44	ARCHON	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
45	DOAR	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
46	SAIL eprints	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
47	Sheetmusic	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
48	ROAR	Yes	No	Yes	Yes	No	Yes	Yes	Yes
49	Celestial	Yes	Yes	No	Yes	Yes	Yes	No	Yes
50	Experimental, UIUC	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
51	Mathematics e print	No	No	Yes	No	Yes	Yes	Yes	Yes
52	Digital commons	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
53	Eprints Archive	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
54	DARE	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
55	Open language	Yes	No	Yes	Yes	Yes	Yes	Yes	No
56	CERN	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
57	TORII	No	Yes	Yes	Yes	Yes	No	Yes	Yes
58	Scirus	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
59	RDN	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
60	CYCLADES	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes

Sometimes, metadata records are not displayed due to some error and these errors can be of numerous types. 49 (81.66%) harvesters show bad argument, 50 (83.33%) bad resumption token, 53 (88.33%) bad verb, 50 (83.33%) can't disseminate format, 51 (85%) ID doesn't exist, 52 (86.66%) no record match, 51 (85%) no ID match, and 49 (81.66%) no set hierarch.

6. Findings

Sixty major metadata harvesting service providers were studied from around the world, eight of which are from India.

The United States is the leading country when it comes to metadata harvesting service providers, followed by the United Kingdom. Among the eight Indian service providers four are disciplinary and the other four are general. These are: Search Digital Libraries, Scientific Journal Publishing in India: Indexing and Online Management (SJPI), Search Engine for Engineering Digital Repositories (SEED), Open J-Gate, Open Index Initiative, Knowledge Harvester, Cross Archive Search Service for Indian Repositories (CASSIR) and Prototype Digital Archive of Indian Aerospace Research (P-DAINAR).

Indian service providers use PKP software for harvesting while Dspace is used by most of the other international harvesters. The majority of the service providers are multidisciplinary. In India Dspace is the most widely used software, followed by eprints. Most of the service providers allow all types of searches like simple search, advanced search, keyword search, author search and subject search.

The majority of the service providers use the Dublin core format for displaying metadata; most do not have an express metadata policy. The metadata harvesting service providers are OAI-compliant and use OAI as metadata prefix support. They use gzip compression for data downloading, they all keep trace of deleted records, and their date granularity form is YYYY-MM-DD.

Metadata harvesting service providers maintain a strong user support system, which helps the user to navigate with ease and retrieve relevant documents. The service providers verify the integrity and authenticity of digital documents by avoiding spoofing (one organisation supplying misleading metadata for a resource belonging to another organisation) and spamming (artificially repeating keywords to boost a page's ranking). A Cross-Archive Service (ARC) is an experimental research service, used to investigate issues

in harvesting OAI-compliant repositories and making them accessible through a unified search interface. It is not a production service and may be subject to unscheduled service interruptions and anomalies.

7. Conclusions

The World Wide Web has created a revolution in the accessibility of information. The development and application of metadata represents a major improvement in the way information can be discovered and used. New technologies, standards, and best practices are continually advancing the applications for metadata. The Open Access movement aims to provide free and open access literature to the scholarly community on the web. In order to be successful in its noble cause, such vehicles must have strong metadata systems. In order to make open access literature globally accessible, Open Access Initiatives worldwide are adopting advanced and developed metadata tools, techniques, standards and softwares to create, preserve and harvest the metadata. A number of metadata harvesting service providers are doing excellent work in harvesting open access vehicles and open access literature scattered on the web.

References

ALCTS/CCS/Committee on Cataloging (2000), *Description and Access. Task Force on Metadata, Final Report*, June 16, http://www.libraries.psu.edu-/tas/jca/ccda/tf-meta6.html (retrieved 25 July 2009).

Baker, T. (2009), DCMI Usage Board review of application profiles, http://dublincore.org/usage/documents/profiles/index.shtml (retrieved 12 July 2009)

Coleman, Nye Pamela (2008), 'Using the Open Archives Initiative Protocol for Metadata Harvesting', *The American Archivist* 71(2), pp. 569–571.

Hodge, Gail (2003), *Metadata Made Simpler*. Bethesda, MD.: NISO Press, http://www.niso.org/news/Metadata_simpler.pdf. (retrieved 5 August 2009).

Munshi, Usha Mujoo (2009), 'Building Subject Gateway in a Shifting Digital World', *DESIDOC Journal of Library & Information Technology* 29(2), p. 9.

Websites Referred to in the Text

UKOLN website on metadata, http://www.ukoln.ac.uk/metadata (retrieved 10 July 2009).